THE UNIVERSITY *of* EDINBURGH

Investigation of DNA Conformation and Enzyme-DNA Systems Using Fluorescence Techniques

Long Ma

Thesis presented for the degree of Doctor of Philosophy
University of Edinburgh
2012

I have no special talents. I am only passionately curious.

—Albert Einstein

Education is an admirable thing, but it is well to remember from time to time that nothing worth knowing can be taught.

—Oscar Wilde

We must accept finite disappointment, but we must never lose infinite hope.

—Martin Luther King, Jr.

Declaration

I declare that this thesis was composed by me, and the research presented is my own except where otherwise stated.

Long Ma
June 2012

Acknowledgements

From the bottom of my heart, I am indebted to a lot of people, who have helped me during the course of my PhD. In the first place, I would like to thank my supervisors Drs. Anita Jones and David Dryden for their trust, help and unceasing encouragements.

Also, I am very grateful to all present and past members of Jones and Dryden's research group, including Drs. Kai Chen, Xiaohua Wu, Gareth Robert, John White Patricia Richardson, Augoustininos Stephanou and Mr. Laurie Cooper and Miss Nisha Kanwar. I am grateful to Kai and Xiaohua for showing me the ropes on experiments and also lots of conversation to keep me active; to Gareth and Laurie for good advices on experiments and also interesting topics during lunch time; to John and Patricia for assistance in DNA synthesis and time-resolved fluorescence experiments respectively. I would like to acknowledge Dr. Jochen Arlt for his great help during the time when I was working in the FLIM (Fluorescence lifetime imaging microscopy) lab in COSMIC (Collaborative Optical Spectroscopy, Micromanipulation and Imaging Centre, University of Edinburgh, UK) and Dr. Scott Cockroft for his help in setting up DNA synthesis suite and nanopore-based single-molecule instruments. Many thanks to my beloved parents for their massive support, care and understanding all the time. Thanks also go to all my friends for their friendship, enjoyable times and exchange of knowledge during my years of study in the University of Edinburgh.

This work was supported by CSC (China Scholarship Council), Edinburgh University International MTEM Scholarship Scheme and EaStCHEM Research School.

Abstract

As a structural analogue of adenine (6-aminopurine), 2-aminopurine (2AP) is a powerful fluorescent probe, when substituted in DNA in place of the natural adenine. Time-resolved fluorescence measurements of 2AP-labeled oligonucleotides, together with steady-state spectroscopy give us an in-depth view of DNA-enzyme interactions, especially the conformational dynamics in solution phase. Herein, this technique has been extended to the study of the transient unzipping of DNA bases, to investigate the structure of three-way junction (3WJ), and the role of base unzipping in the mechanism of human flap endonuclease (FEN).

Seven 2AP-labelled 3WJs were investigated, each containing only one 2AP base in place of adenine. In four of the 3WJs, 2AP was placed in the long duplex region of an arm; while in the other three 3WJs, 2AP was placed near or in the branch point. Comparative time-resolved fluorescence measurements on the 3WJs and corresponding ssDNA and dsDNA controls were made to study the base dynamics, in particular the possibility of unzipping in the vicinity of the branch point. In combination with single-molecule FRET measurements and molecular dynamics simulations, the local and global structure of a DNA 3WJ in solution could be unraveled. It was found to adopt a Y-shaped, pyramidal structure, in which the bases adjacent to the branch point are unzipped, despite the full Watson-Crick complementarity of the molecule.

Human flap endonuclease (hFEN) is divalent metal ion-dependent phosphodiesterase. hFEN carries out structure-specific hydrolysis of 5' bifurcated DNA endonucleolytically. Cleavage occurs at a position one nucleotide into the downstream duplex region. Previous structural, biochemical and modeling studies suggested a double-nucleotide unzipping mechanism at single/double strand junctions for scissile phosphate placement. To confirm this mechanism, 2AP time-resolved fluorescence spectroscopy was used to investigate nucleotide unzipping in hFEN substrates. 2AP was substituted at positions +1 and -1 (relative to the scissile phosphodiester) respectively, in double flap substrates. A series of hFEN

mutants including Y40A, R100A, K93A, were used in this study. In the experiments, ssDNA, dsDNA substrates, DNA substrate-enzyme complexes were investigated in order to elucidate the enzyme-induced distortion of the substrate at the +1 and -1 positions.

TseI is a thermophilic type II restriction enzyme which has ideal activity at an elevated temperature. It is able to recognise and cut the 5 bp palindromic sequence of 5'-GCWGC-3' (W=A or T). A range of biophysical methods have been applied to investigate this enzyme, including size-exclusion chromatography; fluorescence anisotropy (K_d value determination); denaturing HPLC for DNA cleavage analysis on matched and mismatched substrates; fluorescence-based activity assay (K_M, V_{max}, k_{cat}, specificity constant values determination); steady-state fluorescence measurements (DNA-enzyme interaction study). The DNA cleavage characteristics of TseI were fully studied and it was found that it cuts A:A and T:T mismatches in CAG and CTG repeats. This potentially makes it a useful tool for exploring unusual DNA structures containing super-long CAG and CTG repeats which are involved in the aetiology of some neurodegenerative diseases, such as Huntington's disease (HD).

EcoP15I is a type III restriction-modification enzyme whose recognition sequence is 5-CAGCAG-3'. Methyltransferase EcoP15I (M.EcoP15I) adds a methyl group to the second adenine, in the presence of cofactor *S*-adenosyl methionine (SAM). SDS-PAGE, densitometry and size-exclusion HPLC were applied to confirm that EcoP15I adopts a Res_1Mod_2 stoichiometry in solution. The large structural distortion of its substrate (base flipping) by M.EcoP15I was investigated by both steady-state and time-resolved fluorescence. Also, nine 120 mer DNA duplexes, each containing two reversely oriented recognition sites were used to study matched and mismatched sequence cleavage by R.EcoP15 and a cleavage pattern was revealed.

Abbreviation

2AP: 2-aminopurine

3WJ: three-way junction

A: adenine

AAG: alkyladenine glycosylase

A-factor: lifetime fractional amplitude

AFM: atomic force microscopy

AlkA: alkylation glycosylase

ATP: adenosine triphosphate

bp: base pair

BSA: bovine serum albumin

BTT: 5-benzylthio-1H-tetrazole

C: cytosine

CPG: controlled-pore glass

CT: charge transfer

Da: dalton

D-A: fluorescent donor-acceptor

DCM: dichloromethane

DMT: dimethoxytrityl

DNA: deoxyribonucleic acid

dsDNA: double-stranded DNA

DTT: dithiothreitol

E. coli: *Escherichia coli*

EDTA: ethylenediaminetetraacetic acid

E_m: emission wavelength

EM: electron microscopy

ESI-FTICR: electrospray ionization fourier transform ion cyclotron resonance

ETT: 5'-ethylthio-1H-tetrazole

E_x: excitation wavelength

FEN: flap endonucleases

FWHM: full width at half maximum intensity

G: guanine

HD: Huntington's disease

HEPES: (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HEX: hexachlorofluorescein

HPLC: high performance liquid chromatography

IPTG: iso-propylthio-β-D-galactopyranoside

IRF: instrument response function

k_{cat}: rate constant at substrate saturation

K_d: dissociation constant

kDa: kilodalton

K_M: Michaelis-Menten constant

M.W. molecular weight

MD: molecular dynamics

MES: 2-[N-morpholino]ethanesulfonic acid

MFD: multi-parameter fluorescence

min: minute

MS: mass spectroscopy

MTase: methyltransferase

MUG: mismatch-specific uracil glycosylase

nESI: nano-electrospray ionization

NMR: nuclear magnetic resonance

ns: nanosecond

PAGE: polyacrylamide gel electrophoresis

PDB: protein data bank

ps: picosecond

REase: endonuclease

R-M: restriction-modification

RNA: ribonucleic acid

rpm: revolutions per minute

s: second

SAH: *S*-adenosyl-homocysteine

SAM: *S*-adenosyl methionine

SDS: sodium dodecylsulfate

Sf: sinefungin

SM-FRET: single-molecule Förster resonance energy transfer

ssDNA: single-stranded DNA

T: thymine

TCA: trichloroacetic

TCSPC: time-correlated single photon counting technique

TEAA: triethylammonium acetate

T_m: melting temperature

Tris: tris(hydroxymethyl)aminomethane

TRD: target recognition domain

UDG: uracil DNA glycosylase

UV-Vis: ultraviolet-visible

V: volt

WT: wide-type

ε: molar extinction coefficient (molar absorption coefficient, molar absorptivity)

τ: fluorescence lifetime

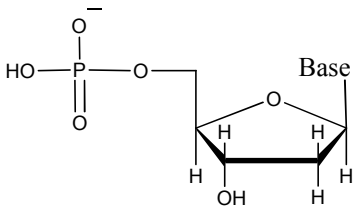
Table of Contents

Chapter 1---Introduction

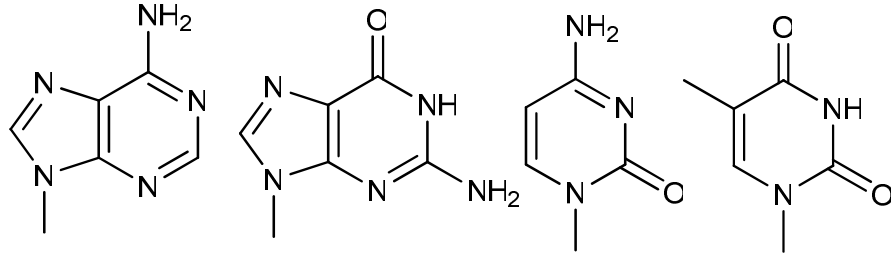
1.1 Biological Context of DNA Structure

DNA (Deoxyribonucleic acid) is a nucleic acid containing the genetic information used in the replication, development and functioning of all known living organisms (with the exception of RNA viruses). The DNA segments carrying this genetic information are called genes. Likewise, other DNA sequences have structural purposes, or are involved in regulating the use of this genetic information. Along with RNA and proteins, DNA is one of the three major macromolecules that are essential for all known forms of life. Proposed by James Watson and Francis Crick[1] in 1953, the double-helical structure determination of DNA is often regarded as the birth of modern molecular biology. DNA is composed of individual repeating units called deoxyribonucleotides, linked by phosphodiester bonds to form linear, unbranched two-stranded polymers. The terminal residue whose C5 is not linked to another nucleotide is called the 5 prime end (5'), and the terminal residue whose C3 is not linked to another nucleotide is called the 3 prime end (3'). DNA polymers can be enormous molecules containing millions of nucleotides and the sequence starts with a 5' phosphate and ends with a 3' hydroxyl. Conventionally the sequence of nucleotide residues in DNA is written, left to right, from the 5' end to the 3' end. The four nucleotides are made up of three different parts: a 2-deoxyribose pentose sugar; a phosphate group; and a nitrogenous base, one of cytosine (C), thymine (T), adenine (A) or guanine (G) as shown in Figure 1.1.

(A) Chemical structure of deoxyribonucleotide



(B) The four bases in DNA



Adenine (A) Guanine (G) Cytosine (C) Thymine (T)

(C) DNA polynucleotide

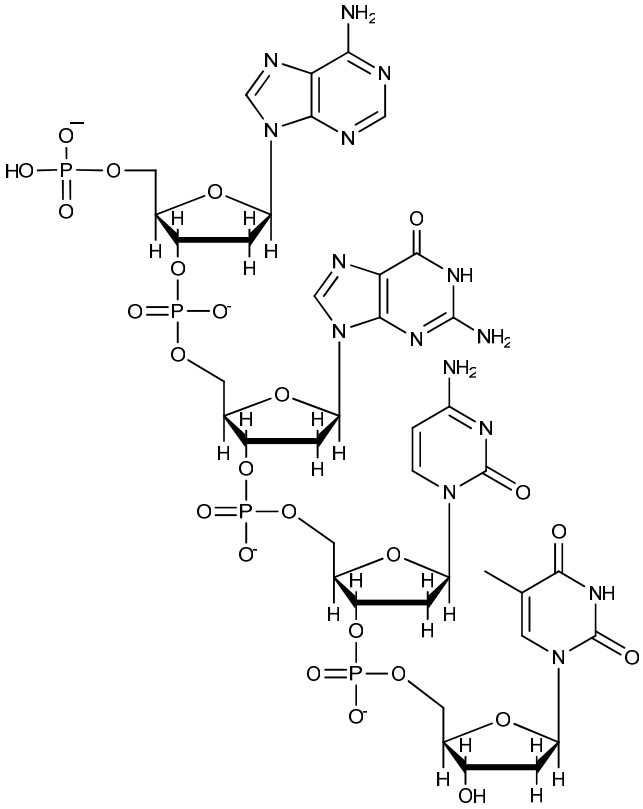


Figure 1.1—Basic chemical structure of DNA. (A) The general structure of a deoxyribonucleotide, found in DNA. (B) Four nitrogenous bases of nucleotides. (C) A short DNA polynucleotide showing the structure of the phosphodiester bond.

The predominant DNA structure found under physiological conditions is B-form DNA, which contains two anti-parallel and mutually complementary stands of nucleic acids connected by A:T and G:C base pairs as shown in Figure 1.2.[1] These two anti-parallel strands spiral around a central axis as shown in Figure 1.3. B-form DNA adopts a right-handed helical structure and each base pair is stacked almost perpendicular to the central axis at a 3.4 Å interval.[2] The structure can be substantially stabilised by the π-π interaction between aromatic rings of each base thus maintaining the duplex nature of DNA. The angle of each base pair plane of B-form double-stranded DNA (dsDNA) is rotated approximately ~34^{o} compared to its neighbouring counterpart, resulting in a complete helical turn for every ~10 continuous base pairs.

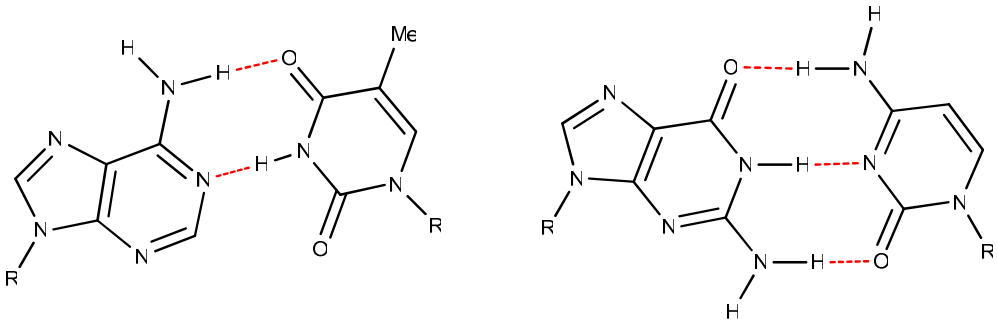


Figure 1.2—Watson-Crick base pairing *via* hydrogen bonds are shown in red between adenine and thymine (left) and guanine and cytosine (right).

The conformation of the DNA molecule is highly dynamic with its base pairs breaking and reforming on the millisecond timescale. The motions of any base relative to its adjacent base can be described using six coordinates: shift, slide, rise, tilt, roll and twist as shown in Figure 1.4.[3,4]

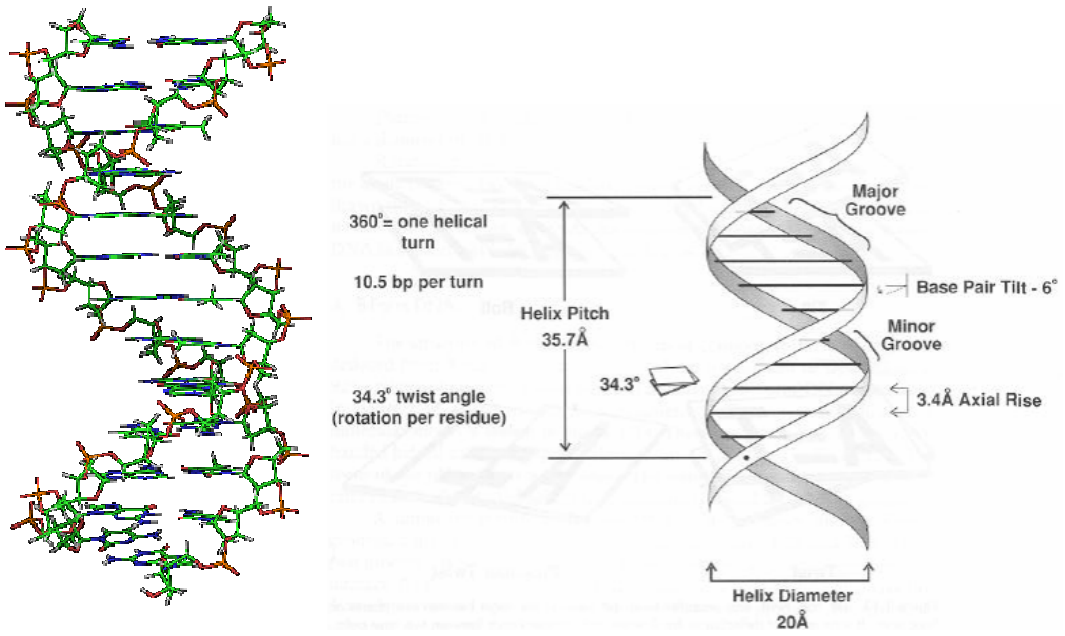


Figure 1.3—Crystal structure of B-form dsDNA (left; PDB: 3BSE) and structural parameters of the B-DNA helix in solution (right).

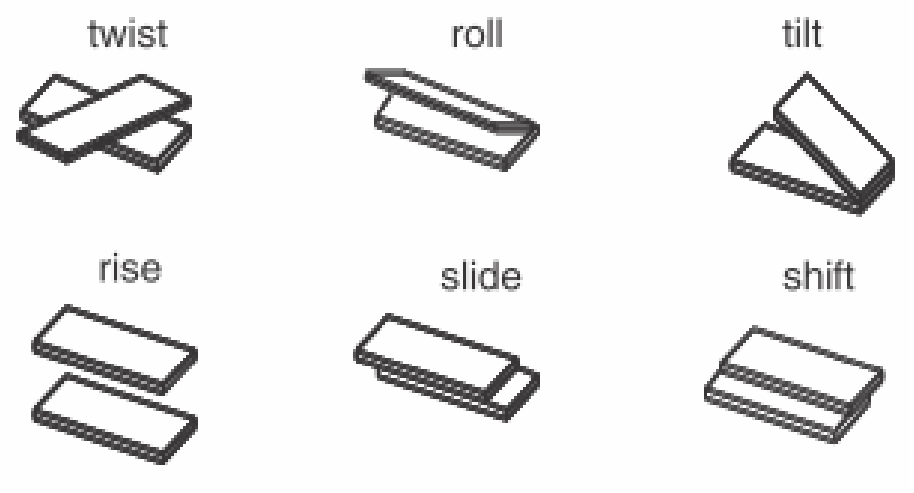


Figure 1.4—The six base pair parameters: twist, roll, tilt, rise, shift and slide. DNA base pairs are shown by rectangular solids (reprinted from Ghorbani and Mohammad-Rafiee[3]).

The canonical right-handed double-helical Watson-Crick model for B-form DNA is the most commonly known DNA structure. However apart from this, other forms of DNA have been observed and it is believed that the DNA molecule can adopt different structures strongly depending on base sequence and external environment as shown in Figure 1.5. The various forms of DNA have been identified as A, B and Z. A comparison of major structural features of A-DNA, B-DNA and Z-DNA are listed below in Table 1.1.[5-8]

	A	**B**	**Z**
Helical sense	Right handed	Right handed	Left handed
Diameter	~26 Å	~20Å	~18Å
Base pairs per helical turn	11	10	12 (6 dimers)
Helical twist per base pair	33°	36°	60° (per dimer)
Helix pitch (rise per turn)	28	34	45
Helix rise per base pair	2.6	3.4	3.7
Base tilt normal to the helix axis	20°	6°	7°
Major groove	Narrow and deep	Wide and deep	Flat
Minor groove	Wide and shallow	Narrow and deep	Narrow and deep
Sugar pucker	C3'-endo	C2'-endo	C2'-endo for pyrimidines; C3'-endo for purines
Glycosidic bond	Anti	Anti	Anti for pyrimidines syn for purines

Table 1.1—The structural parameters of ideal A-, B- and Z-DNA.

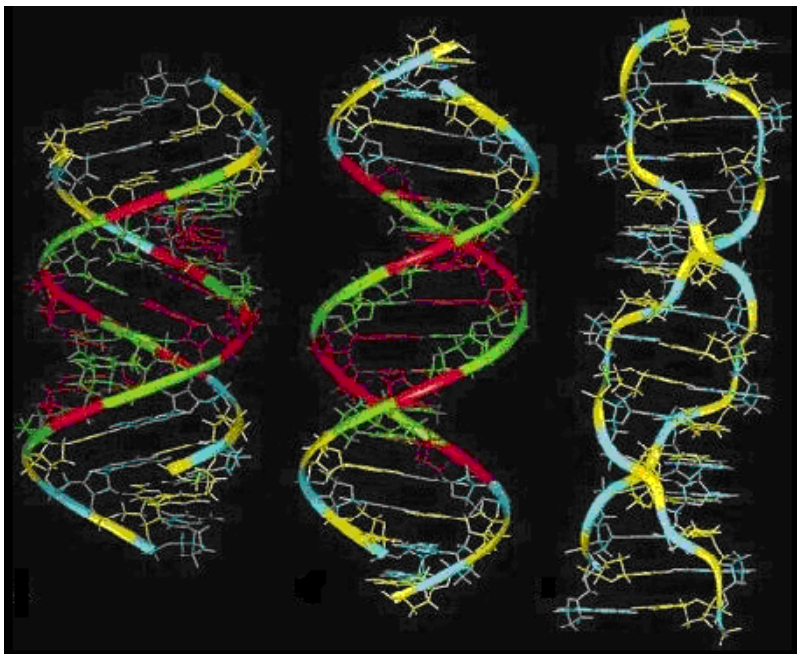


Figure 1.5—Ball-and-stick model structures for A-, B- and Z-DNA (reprinted from Ghosh *et al.*[9]). The nucleotides are coloured respectively: cytosine in yellow, guanine in cyan, thymine in green and adenine in red and a ribbon is superposed depicting the phosphate backbones connecting individual nucleotides. A-DNA (left) and B-DNA (middle) are both displayed as right-handed double-helical structures, while Z-DNA (right) is a left-handed double helix with a dinucleotide repeat and the backbone follows a zigzag path.

1.2 Introduction to 2-Aminopurine

Discovered in 1969 by Ward *et al.*[10], 2-aminopurine (2AP), an analogue of adenine (as shown in Figure 1.6) is one of the most used fluorescent base analogues. Free 2AP has a very high quantum yield in solution, as much as 0.68, while the value for naturally occurring nucleic acid bases is extremely low, making them very difficult to be detected by routine instruments. The high fluorescence quantum yield of 2AP free in solution would be dramatically reduced by ~100 times when incorporated into nucleic acids. This sensitivity to the microenvironment enables it to be a structural probe and it has been utilised in a range of studies such as nucleic acid structures and dynamics, nucleic acid-protein interactions.

2AP, which is one of the commercially available fluorescent base analogues, can base pair stably with thymine and uracil but also moderately with cytosine[11-13]. The lowest energy absorption band of 2AP is centred around 305 nm and has a molar extinction coefficient (ε) of 6000 $M^{-1}cm^{-1}$.[14,15]

The main reason responsible for the considerable decrease in fluorescence intensity of 2AP when it is placed inside DNA is the fast charge transfer (CT) reaction within the picosecond-nanosecond (ps-ns) time domain between the excited 2AP and the neighbouring natural nucleobases mediated by base stacking interaction and base dynamics.[16] Excited 2AP and its charge transfer partner for each of the natural nucleobases have a unique signature for quenching dynamics. Based on the oxidation potentials of the nucleobases in solution, the most efficient charge transfer partner is guanine.[17] The advantages of 2AP when incorporated into nucleic acids can be summarised as:

(1) it can be incorporated into the nucleic acid without obviously perturbing the B-form DNA structure (2) 2AP can be selectively excited in the presence of other natural nucleobases since it can be efficiently excited at wavelengths 305-320 nm. Having a maximum absorption at ~260 nm, natural bases are not be able to absorb the excitation light at this wavelength range (3) the fluorescence of 2AP is strongly quenched within DNA, and this quenching is sensitive to local and global changes in DNA conformation. (4) more importantly, the emission and quenching mechanisms of 2AP have been extensively studied thus the observed fluorescent signal (spectrum or decay) can be reasonably interpreted. This sensitivity to the microenvironment of 2AP when incorporating into DNA has been used in a range of investigations, including nucleic acid structure and dynamics.

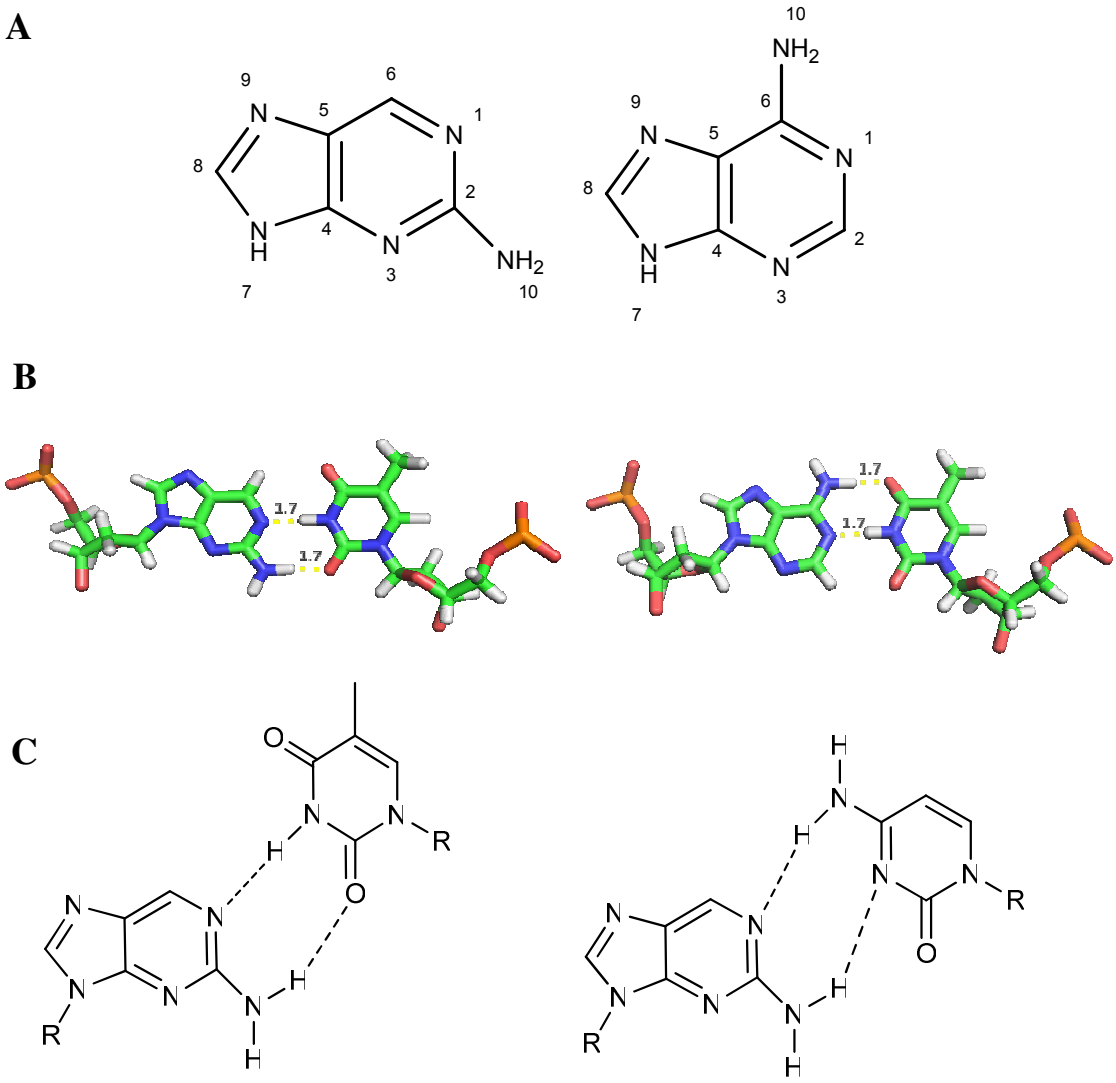


Figure 1.6—Basic properties of 2AP. (A) comparison of the chemical structure of 2AP and adenine, the left one is the chemical structure of 2AP which has an amine group on the 2'-position while the right one shows the chemical structure of adenine which has an amine group on 6'-position. (B) Base pairing of 2AP to thymine (left) and base pairing of adenine to thymine (right) *via* hydrogen bonding. (C) fluorescent base analogue 2AP base paires with thymine (left) and cytosine (right). R=(deoxy)ribose.

Guest *et al.*[18] examined the dynamics of the 2AP:X (where X = T, A, G, and C) mismatches in dsDNA, by measuring 2AP motion in the picosecond time domain using time-resolved fluorescence anisotropy decay. What they found was that among the series of 2AP:X mismatches, the native DNA bases differ in their ability to influence the motion of 2AP. The differences reflect the strength of the base pairing

interactions between 2AP and its base-paired partner X. These interactions are strongest when X = T or C.

Stivers[19] used steady-state fluorescence of 2AP to study the metal ion and base stacking effect in abasic (AB) DNA. Ten 2AP-labelled dsDNA sequences were designed, with 2AP in the opposite position of the absaic site (AB-APopp) or adjacent to the site (AB-APadj) on either strand. A comprehensive study of the fluorescence emission and excitation spectra of these AB duplexes and their corresponding parent duplexes indicates that AB-APopp is significantly less stacked than 2AP in the corresponding normal duplex. In general, AB-APadj on the AB strand is stacked, but AB-APadj on the opposite strand shows destabilised stacking interactions. Also the results show that divalent cation ion binding to the AB duplexes contributes to destabilisation effect of the base stacking interactions in AB-APopp duplexes. However, it has minimal or no effect on the stacking interactions of AB-APadj duplexes. In agreement with these results, the fluorescence of AB-APopp is 18 to 30-fold more sensitive to an externally added quenching agent than the parent normal duplex. When uracil DNA glycosylase binds to AB-APopp duplexes in the presence of 2.5 mM $MgCl_2$, a 3-fold decrease fluorescence can be observed, implying that the unstacked 2-APopp becomes more stacked upon enzyme binding and further suggesting that such conformational changes may occur when this kind of DNA repair enzyme binds AB sites.

To complement steady-state 2AP fluorescence measurements, Rachofsky *et al.*[20] used time-resolved fluorescence of 2AP to characterise the conformational states and dynamics of abasic dsDNA. Their results indicate that divalent cation binding of abasic dsDNA makes the abasic site shift equilibrium between two conformations: a ‘closed’ conformation and an ‘open’ conformation. Because the lifetime and intensity decay kinetics of 2AP incorporated into DNA are sensitive primarily to collisional interactions with the neighboring bases, the absence of dynamic quenching in the open state strongly suggests that the fluorescent base is extrahelical in this conformation. Time-resolved quenching studies reveal that the open state is inclined to quenching by potassium iodide, but the closed state is not. Further results imply

that the conformation of the abasic site varies in a sequence-dependent manner. Undamaged sequences in which the abasic site is replaced by thymine do not exhibit an open state and have different levels of both static and dynamic quenching than their corresponding damaged duplexes. These differences in structure and dynamics could significantly contribute to the highly specific affinity of repair enzymes to the abasic site.

More recently, 2AP has been used as a probe in the rapidly expanding area of small RNA and RNA elements. To test the local conformational changes of *S*-adenosyl- methionine (SAM)-II riboswitch induced by Mg^{2+} and SAM, Haller *et al.*[21] site-specifically labelled 2AP into riboswitch variants and measured the fluorescence intensities. A rapid increase in fluorescence was observed when adding Mg^{2+} for all variants. The addition of SAM resulted in a further but slower increase in fluorescence intensity for all three constructs, which is consistent with a relatively slower, time-dependent response to SAM binding. Such fluorescence intensity changes in 2AP-labelled RNA suggest a base unstacking of 2AP nucleobase upon Mg^{2+} and SAM addition.

A well-established pump-probe laser technique, ultrafast femtosecond fluorescence up-conversion, which enables time-resolved and wavelength-resolved fluorescence measurements was adapted to study RNA conformational dynamics and heterogeneity by Xia *et al.*[22] as shown in Figure 1.7. They measured the decay of 2AP in the RNA molecules and resolved the multiphasic conformational dynamics information from a single decay profile, providing a quantitative ‘snapshot’ on the base stacking interactions. The distinct decay components and their corresponding amplitudes define the heterogeneous conformation of a given RNA molecule and thus the structural and thermodynamic information could be revealed.

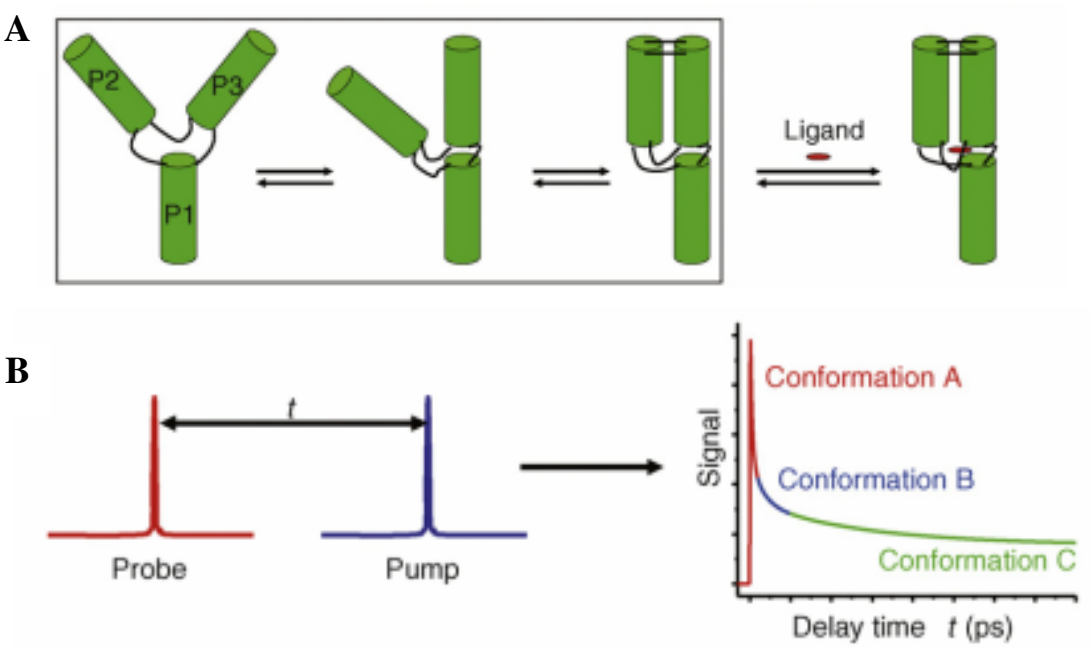


Figure 1.7—Schematic illustration of probing RNA conformational complexity by ultrafast spectroscopy (reprinted from Xia *et al.*[103]). (A) Equilibrium amongst different RNA transient conformations of a three-way junction, where one of them can be stabilised by ligand binding. (B) Ultrafast time-resolved pump-probe approach for probing conformational heterogeneity of RNA. Different conformations can be quantitatively resolved on the basis of their distinct decay dynamics on different timescales (red, blue, and green components).

Experimentally, the phenomenon that the fluorescence intensity of 2AP in DNA is weak compared with free 2AP in solution was adopted by Allan and Reich[23] to investigate the DNA conformation change upon binding of enzyme. What they observed was the binding of M.EcoRI to 2AP-labelled dsDNA, with 2AP at the target position for methylation, brought about a 14-fold increase in the fluorescence intensity with the emission λ_{max} blue shifted, indicating not only that the 2AP was in an extrahelical conformation but also it was positioned into an environment with a significantly lower dielectric constant. Their experiment demonstrates the feasibility of using 2AP-labelled dsDNA as an indicator for establishing a novel fluorescence-based assay to detect DNA conformational alterations especially base flipping within enzyme-DNA complexes. Subsequently, 2AP intensity measurements were carried out on other methyltransferases in solution including M.EcoRV[24], M.HhaI[25], M.TaqI[25], T4 Dam[26], M.EcoRI[23,27], M.KpnI[28], M.EcoP15I[29], M.EcoKI[30] and EcoDam[31].

Additionally, 2AP has also been utilised to study other DNA-protein interactions such the Klenow fragment of DNA polymerase,[32-34] RNA polymerase,[35-37] the restriction endonuclease[38-43] and the uracil DNA glycosylase[44] and helicase[45]. Though the steady-state 2AP measurement is a practical way for monitoring DNA conformation change upon enzyme binding, there is still some uncertainty which can not be clarified by it. Since for the most of the target bases in DNA substrate, they would be put in to a more hydrophobic enzyme binding pocket, and sometimes the direct contact of 2AP with some of the amino acid residues affect the steady-state fluorescence intensity. In this situation, the real fluorescence intensity change caused by base flipping is obscure. Also it is difficult for steady-state measurement to mirror the equilibrium or the population of different base conformations. To solve the abovementioned problems, time-resolved fluorescence measurement of 2AP has been introduced.

Unlike 2AP ribose exhibiting a mono-exponential decay, when it is incorporating into DNA, the lifetime of it is complex and can be best described by four components (τ_1 to τ_4). Different lifetimes are descriptive of different quenching environments corresponding to different conformations of the duplex.[18,20]

The first of the four lifetimes is tens of picoseconds in magnitude. This shortest lifetime (τ_1), corresponds to a conformation of the duplex where the 2AP is in the most intrahelical state highly stacked with its neighbouring bases, the quenching mechanism of this conformation is efficient charge transfer. The longest lifetime, approximately 10 ns, is believed to correspond to the extrahelical states of 2AP thus free from charge transfer, which is comparable to the lifetime of a solvent-exposed 2AP. The two intermediate lifetimes, one around hundreds of picoseconds and the other a few nanoseconds can be attributed to base conformations where the 2AP experiences imperfect stacking with its neighbouring bases thus having more freedom.[46,47] Figure 1.8 provides a simplistic picture of the various base stacking arrangements that would give rise to different lifetimes of the 2AP.

Each lifetime has its own population, termed as fractional amplitude (A-factor). For a well-behaved dsDNA, the shortest components usually account for more than half of the emitting population. The longest lifetime usually constitutes only ~10 percent in abundance. Therefore for a free DNA, the intrahelical conformation is the most dominated one. Note the base conformation in DNA is highly dynamic and is in a fast equilibrium amongst those distinguishable states thus the lifetimes and their fractional amplitudes can be considered as a signature of base dynamics. Each lifetime typifies a distribution of duplex conformations which has similar quenching environments for the 2AP. The essence of this time-resolved lifetime measurements for individual free DNA and DNA-enzyme complexes focus on the changes of the lifetime parameters upon enzyme binding thus the effect of the enzyme manipulation on DNA can be revealed.

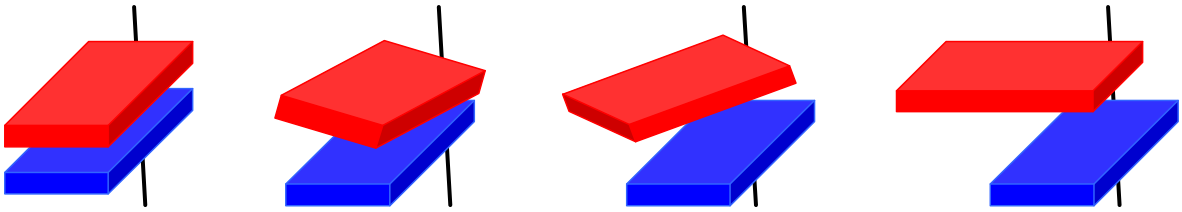


Figure 1.8—An illustration of various stacking conformations between bases (red one represents 2AP base while blue one represents neighbouring natural base) tethered to the phosphate backbone (black), with base-base overlap decreasing from left to right (reprinted from Bonnist[46]).

For example, Neely *et al.*[47] established the time-resolved fluorescence decay function of 2AP in response to base flipping induced by methyltransferase (the principle is shown in Figure 1.9). In order to do so, Neely *et al.*[47] produced a series of DNA-enzyme crystal complexes in which 2AP replaced natural bases in different places in DNA. Four duplexes having 2AP substitution were named in turn: APtarget, in which 2AP replaces the flipped cytosine; APopp, in which 2AP is the base opposite the flipped cytosine; APadj, located immediately adjacent to the flipped cytosine and in the same strand of the duplex; and APout, at a sequence position outside of the enzyme-bound region (served as a control). For APtarget-M.HhaI (T250G)-SAH, the loss of the very shortest (~100 ps) decay

component and the large increase in the amplitude of the longest (~10 ns) component clearly signify base flipping in the crystal. The APout-M.HhaI-SAH lifetime parameters are characteristic of 2AP in unbound DNA, having the short τ_1 lifetime in 64% of the complexes. This is consistent with the fact that the 2AP in this complex is intrahelical and well stacked in the duplex. The decay parameters for APopp-M.HhaI-SAH and APadj-M.HhaI-SAH also imply that the majority of the 2AP in these sequences remains intrahelical. All these results are in line with the fact that in the crystal structures of these complexes , 2AP replacing the cytosine in the APtarget-M.HhaI (T250G)-SAH complex is flipped, while 2AP in 2APadj, 2APopp and 2APout all remain stacked in the DNA interior upon enzyme binding. More significantly, they have shown that the same clear response for base flipping for APtarget-M.HhaI (T250G) binary complexes with DNA in solution, giving a definitive signal for base flipping. This can be best evidenced by the fact that upon enzyme binding, the fractional amplitude of the shortest lifetime component (τ_1) dropped from 0.53 to 0.19, while the fractional amplitude of the longest lifetime component (τ_4) increase from 0.06 to 0.46 thus becoming the most dominant one. Almost half of the population of 2AP base in APtarget-M.HhaI (T250G) complex are in a flipped-out conformation.

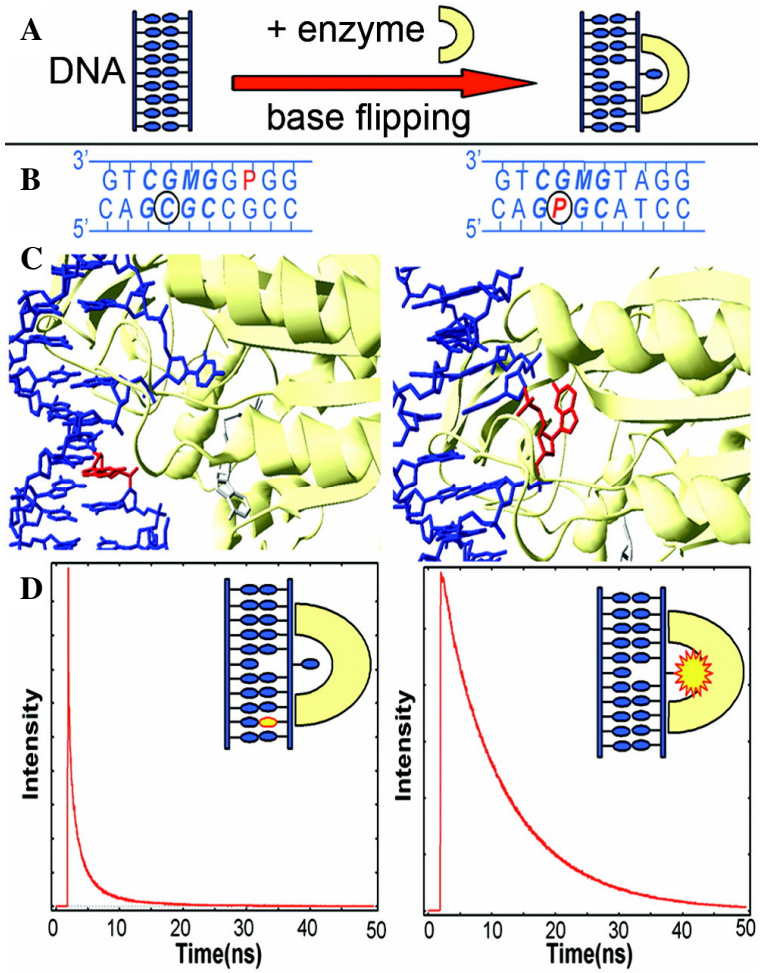


Figure 1.9—Illustration of using time resolved fluorescence of 2AP to study base flipping induced by methyltransferase (reprinted from Neely *et al.*[47]). (A) A schematic representation of base flipping upon methyltransferase binding. (B) The sequences of the 10 bp at the centre of the APout duplex (left) and APtarget duplex (right). Bases in the M.HhaI recognition sequence are shown in bold/italic; the target base is circled; 2AP base is denoted P and is in red; M is 5-methyl cytosine base (used to direct enzyme binding to the opposite strand of the duplex). (C) Crystal structures of the complexes of the M.HhaI enzyme with the APout duplex (left) and the APtarget duplex (right), showing the molecular structure in the vicinity of the recognition sequence. (D) Fluorescence decay response of the crystalline complexes of M.HhaI with APout (left) and APtarget (right).

Taking advantage of lifetime measurement, Neely *et al.*[41] also used time-resolved fluorescence spectroscopy of 2AP-labelled DNA to study the base flipping mechanism of a range of restriction enzymes Ecl18kI, PspGI and EcoRII-C, which are specific for interrupted 5 bp target sequences, flip the central base of their respective sequences into the protein pockets for facilitating sequence recognition

and adjusting the DNA cleavage pattern. Unique lifetime parameters are capable of giving a detailed picture of the molecular microenvironment of the extrahelical flipped-out 2AP base and revealing the heterogeneity, especially the interaction of it with amino acid residues in close vicinity. For instance, Ecl18kI as a base flipper, is able to flip the central base within its recognition site. Surprisingly the decay parameters for Ecl18kI-T/2-Ca^{2+} (T/2 is the DNA sequence in the recognition sequence of Ecl18kI with 2AP in the central position and base paired with T) do not conform to the fluorescence lifetime pattern for base flipping. When mutating the tryptophan residue (W61) within the Ecl18kI flipping pocket to W61A (alanine), the lifetime parameters for the Ecl18kI (W61A)-T/2-Ca^{2+} give rise to a general increases in lifetimes (except for the longest component) also the amplitude of the shortest lifetime component decreases noticeably within this mutant. This signifies a transfer of population from highly quenched to relatively unquenched conformations in the W61A-2/2-Ca^{2+} complex. The co-crystal structure of Ecl18kI in complex with its natural substrate shows the flipped base directly stacked with tryptophan (W61) in the binding pocket. 2AP fluorescence is known to be subject to rapid quenching by electron transfer from tryptophan, taking place on a picoseconds timescale, a rate similar with electron transfer quenching of intrahelical 2AP by guanine. Thus, the observation of the very short (70 ps) decay time and the highly populated conformation on this lifetime for 2AP in the wild-type (WT) Ecl18kI ternary complexes in the solution is entirely consistent with a base-flipped conformation equivalent to that observed for the natural substrate in the crystalline complex. The change of Ecl18kI (W61A)-T/2-Ca^{2+} lifetime parameters compared with Ecl18kI (W61)-T/2-Ca^{2+}, clearly demonstrates the interaction of tryptophan with 2AP. Together with the fact that X-ray structure shows base flipping for Ecl18kI, it can constitute convincing evidence that WT Ecl18kI can flip the base, which can be revealed by time-resolved fluorescence of 2AP. The close similarity between the decay parameters of the Ecl18kI (W61)-T/2-Ca^{2+} to free duplexes suggests that in the WT Ecl18kI, base flipping of 2AP is an efficient process and the binding pocket of Ecl8kI forms a similar environment to the DNA duplex in order to stabilize the extrahelical base. Across all of the lifetime data for R-M base flipping systems, it is clear that the lifetime signature for enzyme flipped-out DNA complex is unique and

evidence for base flipping. Additionally, the reported subtle lifetime changes in enzyme bound 2AP-labelled DNA (with 2AP outside the target sequence) can be attributed to the minimal base disturbance upon enzyme binding. By combining the co-crystal structures, the solution and crystalline phase 2AP lifetime measurements, Neely and co-workers not only linked the solution-base lifetime measurements for base flipping to co-crystal structure but also delineated the dynamics and the microenvironment of the flipped-out base.

1.3 Introduction of Restriction-Modification (R-M) Systems

1.3.1 General overview of R-M systems

The phenomenon of restriction-modification (R-M) systems was first observed genetically and described in the years 1952-1953 by Luria and Human[48], Bertani and Weigle[49]. In their papers, they observed that different kinds of bacteriophages showed different levels of viability to grow in different host strains. Once growth of the bacteriophage had taken place in one strain, the phages could continue to grow happily in this strain while they were restricted in their ability to grow in other strains. They attributed it as host-induced or host-controlled variation, rather than rationalizing the underlying mechanism of R-M systems. In the year 1962, Arber and Dussoix[50] described that the lambda DNA was rejected and degraded upon infection of different bacterial hosts, unless it carried host-specific modification of that DNA, thus laying the foundations for the phenomenon of R-M.

R-M systems function as immune systems adopted by prokaryotes that prevent foreign DNA entering the cells. Typically, R-M systems are composed of two enzyme components responsible for separate functions: a restriction endonuclease (REase) which can recognise a specific DNA sequence and subsequently cleave it to harmless fragments and a cognate methyltransferase (MTase), which is capable of protecting the DNA from being cleaved by modification methylation of adenosyl or cytosyl residues within the sequences recognised by the methyltransferase.

In R-M systems, REases recognise 'non-self' DNA injected into the cell from the phage and plasmids since this DNA lacks characteristic modification within specific recognition sites.[51] Foreign DNA is then inactivated by endonucleolytic cleavage. A question can be raised here for R-M systems as to how they distinguish 'self' and 'non-self' DNA. It would appear host 'self' DNA is normally methylated by the MTase following replication, whereas invading non-self DNA is not. However, it is possible for phage and some plasmids to acquire host modification and to 'escape' restriction, suggesting that the R-M barrier is not completely perfect. Figure 1.10 shows how the R-M systems work as exemplified by the type I R-M system.

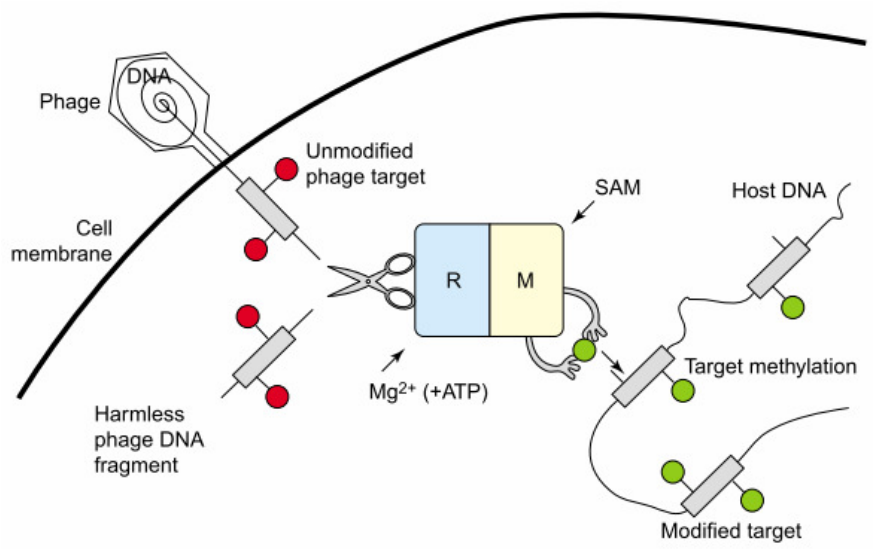


Figure 1.10—The mechanism of R-M systems, as exemplified by type I R-M enzymes (reprinted from Tock and Dryden[52]). These enzymes recognise the methylation state of their specific target sequence. Fully methylated DNA as shown by two green circles on the target sequence on the host DNA can be recognized as part of the bacterial genome. Hemimethylated DNA as shown by a single green circle on the host DNA target sequence can be recognised as newly replicated bacterial DNA, and the MTase (M) modifies the other strand by methylation in the presence of the cofactor SAM. However, foreign DNA, from a phage or plasmid genome, generally doesn't carry specific modification (red circles on the target sequence of phage DNA) and is recognised to be 'non-self' by the REase (R) and cleaved into harmless fragments.

R-M enzymes can cleave dsDNA at the recognition sequence or at a random position away from the recognition sequence. This can be conducted by hydrolysis of one phosphodiester bond in the backbone of each DNA strand as shown in Figure 1.11. REases activity usually requires Mg^{2+} or a similar divalent cation; and some also require adenosine triphosphate (ATP) or SAM. Cleavage generally occurs on the 5'-side of the phosphate group, leaving DNA fragments with 5'-phosphoryl and 3'-hydroxyl termini. Unlike DNA with 3'-phosphoryl and 5'-hydroxyl ends, 5'-phosphoryl and 3'-hydroxyl termini can be rejoined again by DNA ligase.[53]

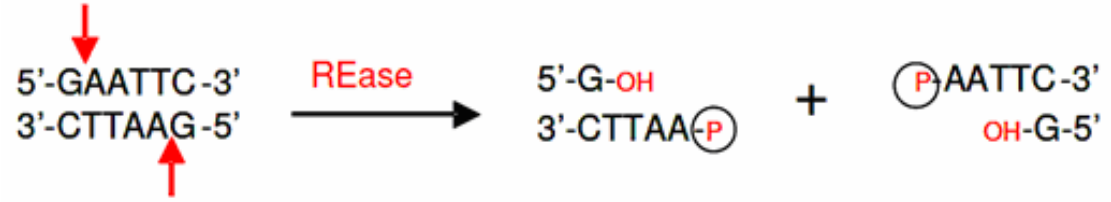


Figure 1.11—An illustration of restriction enzyme cleavage on dsDNA. The red arrow indicates the cutting position. When cleavage takes place, the phosphodiester bond of DNA will be broken, leaving two sticky ends (for some restriction reactions the cleavage can result in blunt ends rather than sticky ends) with 5'-phosphoryl and 3'-hydroxyl termini.

Modification MTase requires SAM as a methyl group donor to supply a methyl group to one nucleotide in each strand of the recognition sequence. Thus, SAM is an essential cofactor for methylation. Adenine and cytosine are the only bases known to be methylated by R-M systems.[54] Adenine can only be methylated if the adenine occurs at the N6 position, while cytosine can be methylated at the N4 or C5 position[53] as shown in Figure 1.12.

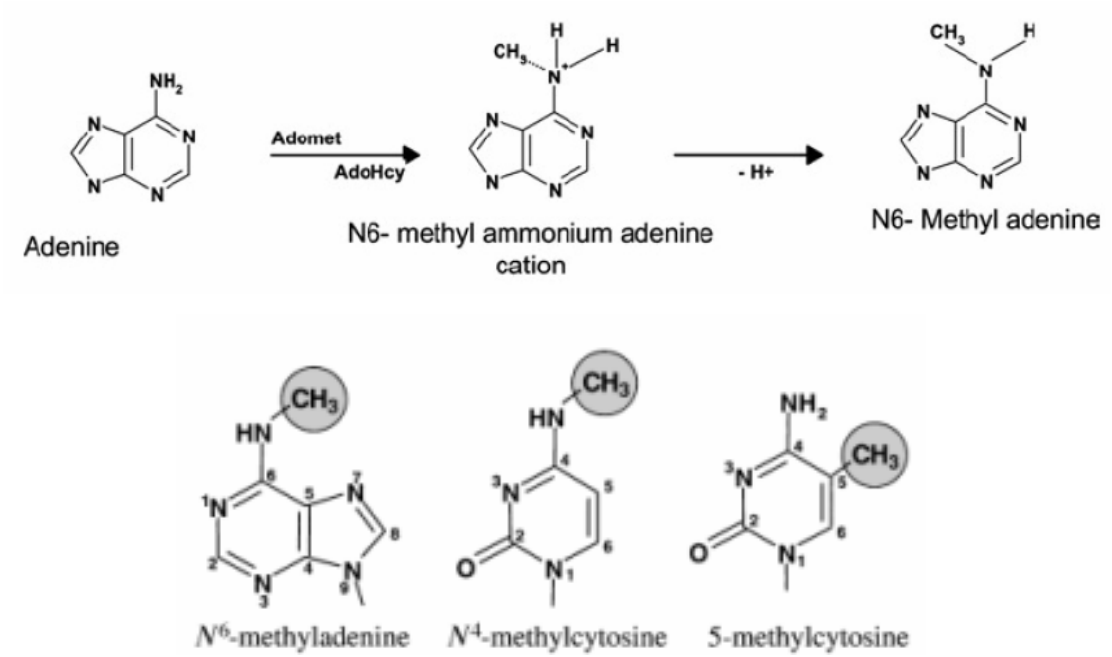


Figure 1.12—DNA methylation by R-M systems. An example of the reaction mechanism for base methylation illustrated using N6-adenine methylation (upper panel, reprinted from Bheemanaik *et al.*[28]). AdoMet is also called SAM; AdoHcy is known as SAH (*S*-adenosyl-homocysteine). Structures of methylated bases occurring in DNA (lower panel, reprinted from Jeltsch[54]).

Classification of R-M systems is based upon their enzyme composition, cofactor requirements, the nature of their target sequence, and the position of the site of DNA cleavage with respect to the target recognition sequence. Four different types of R-M systems have been characterised thus far and have been classified as type I, II, III and IV.[52] Some examples are shown in Table 1.2.

Up until 01 February 2012, REBASE[55,56] (http://rebase.neb.com/rebase/rebase.html) lists 4005 restriction enzymes (97 type I, 3882 type II, 14 type III, 12 type IV) and 996 methyltransferases (68 type I, 791 type II, 20 type III, 117 orphan methylases).

Table 1

Characteristics and organization of the genetic determinants of different classes of R-M systems[a].

	Type I	Type II	Type III	Type IV
Example R-M system	EcoKI	EcoRI	EcoP1I	EcoMcrBC
Genes	*hsdR*, *hsdM*, *hsdS*	*ecorIR*, *ecorIM*	*mod*, *res*	*mcrB*, *mcrC*
Subunits	Three different subunits (R, M and S) combine to form $R_2M_2S_1$ and M_2S_1	Two different subunits (R and M) combine to form R_2 or M_1	Two different subunits (mod and res) combine to form mod_2res_2	Two different subunits are present, McrB and McrC
Enzyme activities	REase, MTase and ATPase	REase or MTase	REase, MTase and ATPase	REase and GTPase
Co-factors required for DNA cleavage	ATP, SAM, Mg^{2+}	Mg^{2+}	ATP, Mg^{2+} (SAM)	GTP, Mg^{2+}
Co-factors required for methylation	SAM	SAM	SAM	No methylation
Recognition sequence	Asymmetric and bipartite, e.g. EcoKI, 5′AAC(N_6)GTGC	Mostly symmetric, e.g. EcoRI, 5′GAATTC	Asymmetric, e.g. EcoP1I, 5′AGACC	Bipartite and methylated, e.g. EcoMcrBC, 5′RmC($N_{30–4000}$)RmC
Cleavage site	Variable locations 1000 bp from recognition site	Fixed location at or near the recognition site	Fixed location 25–27 bp from recognition site	Between methylated bases at multiple sites
DNA translocation	Yes	No	Yes	Yes

Table 1.2—Lists of four types of R-M systems and their specific properties (reprinted from Tock and Dryden[52]).

1.3.2 Type I R-M systems

Type I R-M systems are hetero-oligomeric, multifunctional enzymes that perform both restriction and modification activity. These enzymes recognise asymmetric bipartite nucleotide sequences comprising two components, one of which is 3 or 4 bp and the other of which is 4 or 5 bp, separated by a nonspecific spacer from 6 to 8 bp. All known type I methyltransferase enzymes are N6 adenine-specific DNA methylases, which are able to catalyse the transfer of the methyl group from SAM to the N6 position of specific adenine residues in their respective target sequences. EcoKI for example, with a target recognition sequence of AAC(N6)GTGC, the underlined A is a methylation site, whereas the underlined T represents the site of methylation for adenine on the complementary DNA strand.

The activity of type I R-M enzymes as an endonuclease or a methyltransferase is determined by the methylation state of the target sequence when binding to its target sequence. If DNA is fully methylated, neither methyltransferase nor restriction will take place. If the DNA is hemimethylated, as is the case for newly replicated DNA,

the enzyme methylates the unmethylated strand, resulting in methylated DNA[57]. If the DNA is unmodified, the enzyme, while bound to its target site, is believed to translocate or pull the DNA towards itself simultaneously in both directions in an ATP-dependent manner. This translocation process brings together enzymes bound to different sites on the same DNA molecule, and it is thought that DNA cleavage occurs when this translocation is hindered, *via* either by collision with another translocating complex or by the topology of the DNA substrate.[58]

A type I restriction enzyme is composed of three subunits HsdM (~60 kDa), HsdS (50-60 kDa) and HsdR (~140 kDa) encoded by three closely linked genes, *hsdR*, *hsdM*, and *hsdS*.[59,60] *hsdM* and *hsdS* are transcribed using the same promoter, while *hsdR* is transcribed from its own promoter. The acronym Hsd denotes 'host specificity of DNA'.

The two subunits encoded by *hsdM* and *hsdS*, HsdM and HsdS (often referred to as M and S) are both required for methyltransferase activity. The S (specificity) subunit contains two target recognition domains (TRDs) that confer the target sequence specificity to both the restriction and modification activities of the complex. HsdM includes the binding site for SAM as well as the active site for DNA methylation. The third subunit (HsdR or R) has endonuclease activity, including the active site for ATP hydrolysis and sequences essential for DNA translocation. Figure 1.13 shows the model proposed for the structure of EcoKI bound to DNA.[61,62] Recently Kennaway *et al*.[63] showed structures of two type I R-M enzymes, EcoKI and EcoR124I, derived using electron microscopy (EM), small-angle scattering, and detailed molecular modeling as shown in Figure 1.14.

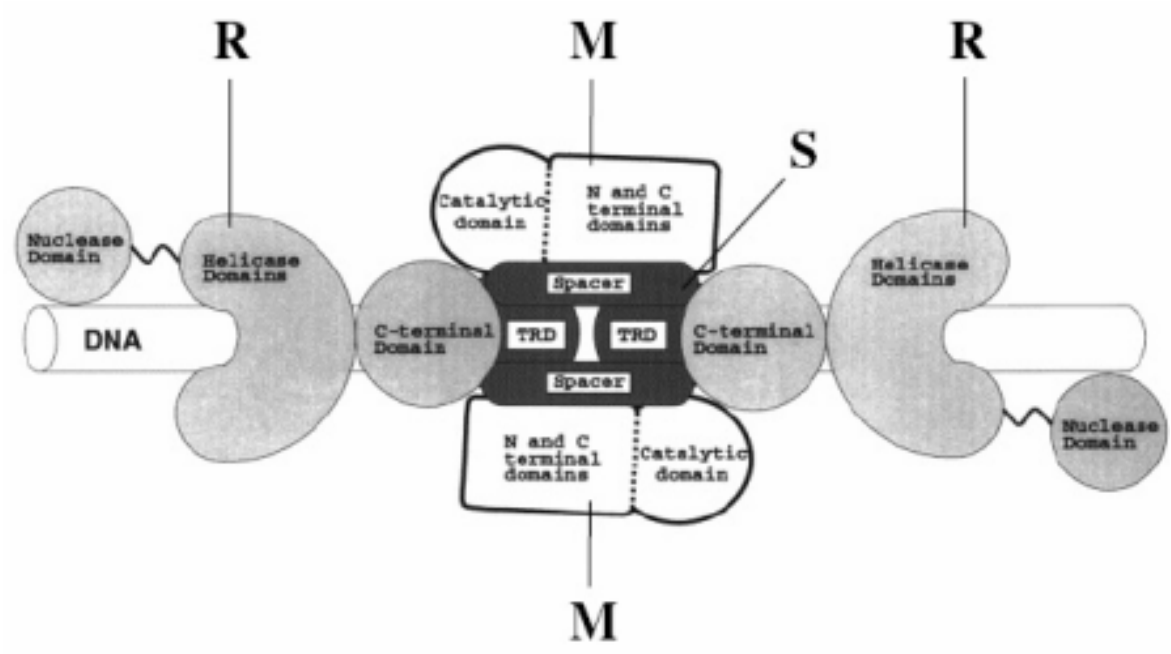


Figure 1.13—A model of the domain structure of a type I restriction enzyme (reprinted from Davies *et al.*[61]). The S subunit (black) comprises two target recognition domains (TRDs) which recognise and bind to specific target sites on DNA (white cylinder). The TRDs are joined by two spacer regions. The M subunits (bold outline, white) are made up of a catalytic domain and N- and C-terminal domains. The N-terminal domain is involved in recognition (of the DNA methylation state) and methylation, whereas the C-terminal domain is involved in the assembly of the S subunit. The R subunits (light grey) comprise an endonuclease domain, responsible for cleavage, a helicase domain, and a C-terminal domain involved in the association with the M and S subunits.

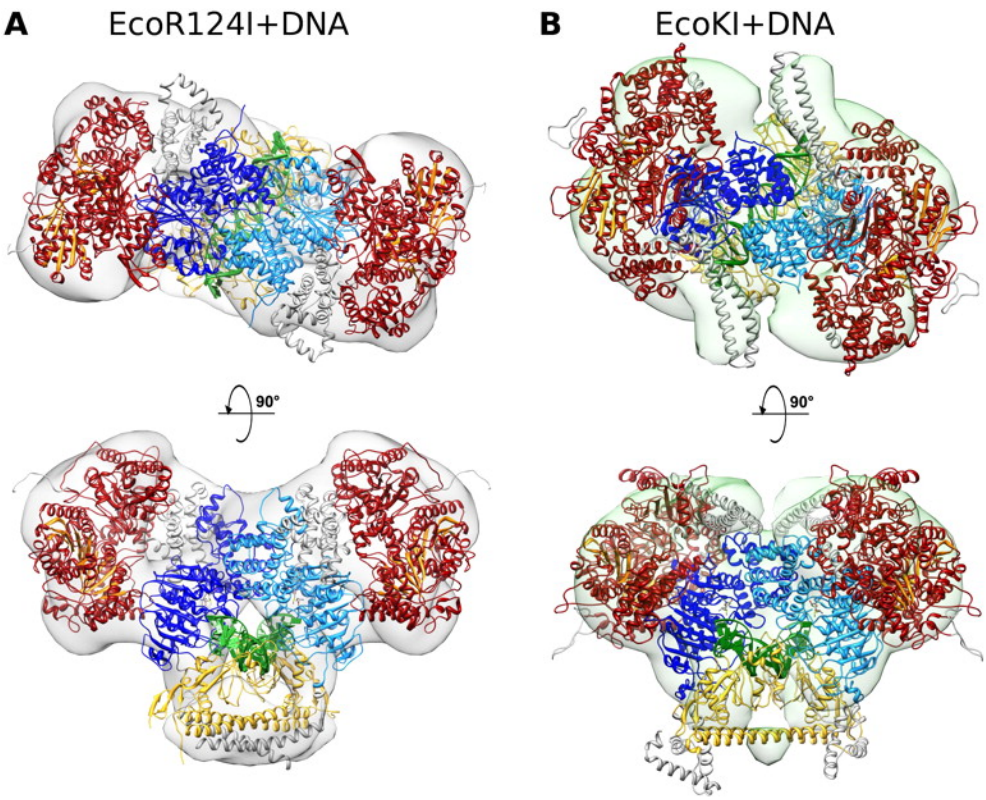


Figure 1.14—Atomic models of EcoR124I+DNA (A), EcoKI in complex with DNA (B) docked into the EM map densities. (A) Two views of the EcoR124I+DNA model showing the MTase core closed around DNA (green). Adenine bases are flipped out into the active sites of each of the two HsdM (light and dark blue), induced by an ~45° bend in the DNA. The HsdS is in yellow, and the two HsdR are shown in red, with the β sheets of the recA-like motor domains coloured orange. Residues missing from the crystal structures (the 44 and 152 C-terminal residues of HsdM and HsdR, respectively) were modeled de novo and are shown in gray. (B) A model for the second type I R-M enzyme, EcoKI bound to DNA. Colours are as in A, with residues modeled de novo shown in gray. The HsdS and HsdM from the MTase structure (PDB: 2y2C) were docked in as a single rigid body. The HsdR modeled on those from EcoR124I (PDB: 2w00), were placed in a position analogous to the EcoR124I model.

1.3.3 Type II R-M systems

1.3.3.1 Diversity

When defining a type II restriction enzyme, the main criterion is if its cleavage site is specifically within or in a close vicinity of its recognition site. For an orthodox type II restriction enzyme, it is usually active as a homodimer in aqueous solution and recognises and cleaves a palindromic sequence 4-8 bases in length. The cofactor required for cleavage is Mg^{2+} and some other similar divalent cations. It cleaves the dsDNA leaving 5'-phosphate and 3'-hydroxyl group. The cofactor acting as a methyl group donor required by MTase is SAM. Methylation of adenines occurs at the N6 position while cytosines are methylated at the N4 or C5 position.[64] Nevertheless, a few enzymes of the type II R-M system can not strictly conform to this definition. Therefore a subdivision of type II restriction enzymes can be proposed as shown in Table 1.3. Most of the restriction enzymes used for recombinant DNA work belong to this subtype, which is called type IIP (P stands for palindromic).

Subtype	Characteristic feature	Example[a]	Recognition sequence[b]
Orthodox	Palindromic recognition site, which is recognized by a homodimeric enzyme, cleavage occurs within or adjacent to the recognition site	***Eco*RI**	G↓ A A T T C C T T A A↑ G
		***Eco*RV**	G A T↓ A T C C T A↑ T A G
		***Bgl*I**	G C C N N N N↓ N G G C C G G N↑ N N N N C C G
Type IIS	Asymmetric recognition site with cleavage occuring at a defined distance	***Fok*I**	G G A T G N_9↓ N N N N C C T A C N_9 N N N N↑
Type IIE	Two sites required for cleavage, one serving as allosteric effector	***Nae*I**	G C G↓ C G C C G C↑ G C G
Type IIF	Two sites required for cleavage, both sites are cleaved in a concerted reaction by a homotetrameric enzyme	***Ngo*MIV**	G↓ C C G G C C G G C C↑ G
Type IIT	Different subunits with restriction and modification activity	*Bpu*10I	C C↓ T N A G C G G A N T↑ C G
		*Bsl*I	C C N N N N N↓ N N G G G G N N↑ N N N N N C C
Type IIG	One polypeptide chain with restriction and modification activity	*Eco*57I	C T G A A G N_{14} N N↓ G A C T T C N_{14}↑ N N
Type IIB	Cleavage on both sides of the recognition site	*Bcg*I	N N↓ N_{10} C G A N_6 T G C N_{10} N N↓ ↑N N N_{10} G C T N_6 A C G N_{10}↑ N N
		*Bpl*I	N N_4↓ N_8 G A G N_5 C T C N_8 N_4 N↓ ↑N N_4 N_8 C T C N_5 G A G N_8↑ N_4 N
Type IIM	Methylated recognition site	*Dpn*I	G ^{m}A↓ T C C T↑ ^{m}A G

Table 1.3—Nomenclature and classification of type II restriction endonucleases (reprinted from Pingoud *et al.*[65]).

Type IIS, type IIE and type IIF restriction endonucleases share some striking similarities in spite of some differences. All these enzymes require two copies of a particular DNA recognition sequence to cleave the DNA. The disparity of these three types with regard to their DNA recognition and cleavage properties are expanded in the following paragraphs.

Type IIS restriction endonucleases such as FokI recognize asymmetric sequences and cleave these sequences at a defined distance.[66] It was believed that these enzymes function as monomers in solution. However, Bitinaite *et al*.[67] provided evidence that FokI, which is a type IIS restriction endonuclease composed of a DNA recognition domain and a catalytic domain, must dimerize *via* the catalytic domain when bound to DNA to be active for the cleavage. FokI requires two copies of its asymmetric recognition sequence and the cleavage occurs at a defined distance downstream of one recognition site.

Type IIE restriction endonucleases interact with two copies of their recognition sequence as well, one being the DNA substrate for cleavage, the other serving as an allosteric effector.[68] For instance, Yang *et al*.[69] showed that NaeI has a DNA concentration-dependent catalysis with sigmoidal kinetics, indicating cooperative and allosteric interaction within the NaeI dimer. EcoRII also belongs to this subtype.[70] The cleavage for this kind of enzyme happens within one of the two recognition sites.

Type IIF restriction endonucleases resemble type IIE enzymes, in as much as they interact with two copies of their recognition sequence. They differ from the type IIE enzymes in that they cleave both sequences in a concerted reaction.[71] Wentzell *et al*.[72] conducted the kinetic investigation of type IIF enzymes on supercoiled plasmids with one or two recognition sites for SfiI. What they found was that it concertedly cleaved at both recognition sites upon binding of two of their DNA recognition sites. Enzymes NgoMIVd[73], Cfr10I[74] and Bse634I[75] also belong to this subdivision. Thus, the active complex comprises a tetrameric enzyme and two recognition sites. A detailed comparison of orthodox type II (type IIP) type IIE, type IIF and type IIS is

shown in Table 1.4.

	Orthodox Type II	Type IIE	Type IIF	Type IIS
DNA substrate requirements	Recognize palindromic DNA recognition sequences	Two copies of their usually palindromic recognition sequence	Two copies of their usually palindromic recognition sequence	Two copies of their asymmetric recognition sequence
Oligomeric status in solution	Dimer	Dimer	Tetramer	Monomer, but active as dimer
Domains per monomer	One	Two, an endonuclease-like catalytic domain and a DNA binding domain	One	Two, an endonuclease-like catalytic domain and a DNA binding domain
Cleavage	Within a single recognition site	Within one of the two recognition sites	Concertedly within or near both recognition sites	At a defined distance downstream of one recognition site
Prototypes	BamHI, EcoRI	EcoRII, NaeI	Cfr10I, SfiI	FokI[a]

Table 1.4—Characteristics of orthodox type IIP REases and of the subtypes IIE, IIF and IIS (reprinted from Mücke *et al.*[76]).

Type IIT restriction endonucleases are composed of two different subunits with restriction and modification activity, such as Bpu10I and BsII. Bpu10I recognizes an asymmetric sequence and functions as a heterodimer (αβ) in which both subunits presumably have one active site.[77] BslI recognizes a palindromic sequence and functions as a heterotetramer ($\alpha_2\beta_2$).[78]

Type IIB restriction endonucleases cleave DNA at both sides of the recognition sequence, for example BcgI which recognizes an asymmetric sequence, or BpII which recognizes a symmetric interrupted sequence. These enzymes are composed of different subunits and are bifunctional enzymes with both DNA cleavage and methyltransferase activities. They require the presence of SAM for DNA cleavage.[79] For BcgI, it was shown that the catalytic centers for restriction and modification are located in the α-subunit, whereas the β-subunit harbors the target recognition domain.[80]

Similarly to type IIB enzymes, type IIG restriction endonucleases are stimulated by AdoMet but have both restriction and modification functions present in a single polypeptide chain rather than different subunits; e.g. Eco57I.[81] Type IIM restriction endonucleases recognize and cleave methylated DNA, for example DpnI.[82]

Some type II restriction enzymes hydrolyse only one strand of the duplex thereby producing DNA molecules that are “nicked”, rather than cleaved. For example,

Nt.BstNBI is a naturally occurring thermostable nicking endonuclease cloned from *Bacillus Stereothermophilus*.[83,84] Nt stands for Nicking enzyme with top strand cleavage activity. These conventional nicks (3'-hydroxyl, 5'-phosphate) can serve as initiation points for a variety of further enzymatic reactions such as replacement DNA synthesis, strand-displacement amplification[85], exonucleolytic degradation or the creation of small gaps.[86]

The type II restriction enzymes comprise a very large family. However, as can be seen, overlap cannot be avoided due to great diversity among them. For example, FokI, the archetypal IIS enzyme, could also be considered as a type IIE enzyme, since it requires binding to a second recognition sequence; as shown by Vanamee *et al.*[87] Undoubtedly, new discoveries of type II restriction enzymes will eventually give rise to new subtypes.

As mentioned above, REases of types IIE, IIF and IIS, in contrast to orthodox type IIP REases, have the capacity for interacting with two DNA recognition sites, whose mechanisms are intriguing. A mechanistic model has been proposed for type IIE REases based on known biochemical and structural data. A comparison of this model with those proposed for type IIF and IIS REases is pictured below in Figure 1.15.

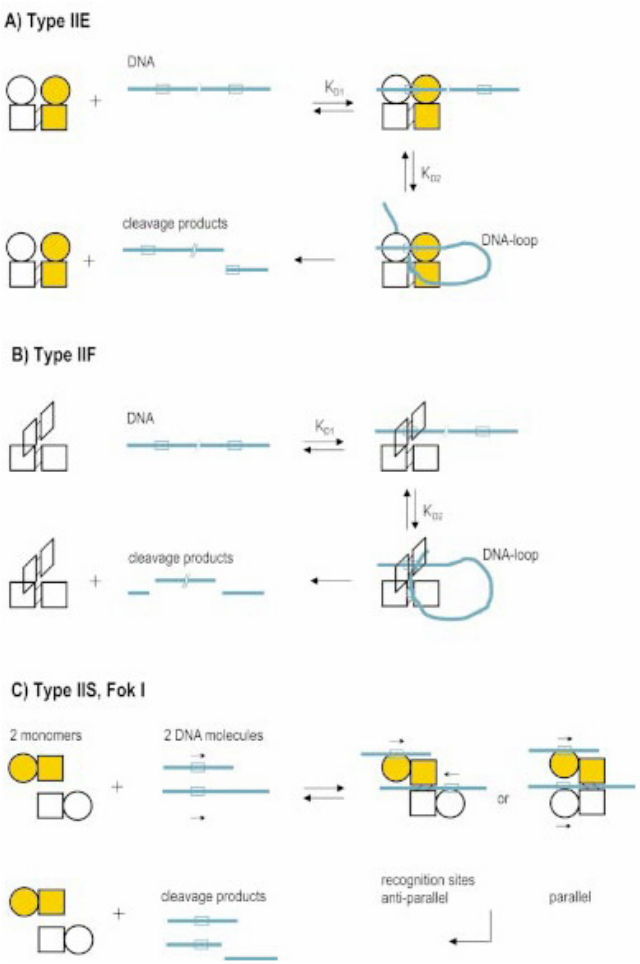


Figure 1.15—Proposed mechanistic models for three type II REase interacting with their respective DNA substrate (reprinted from Mücke *et al.*[76]). (A) Type IIE REase interaction with a DNA molecule containing two sites in cis. The two different domains of a monomer are represented by a circle and a rectangle respectively. The white one and the yellow one represent two monomers forming an active homodimer. The circle represents the DNA-binding domain, while the rectangle represents the endonuclease-like domain. The lines that link both rectangles symbolize dimerization. Blue stretches with boxes denotes DNA with two copies of the recognition sites. The enzyme forms intermediate DNA loops with DNA containing two recognition sites. The DNA will be cleaved at one recognition site, which is bound to the endonuclease domain. (B) Type IIF REase interaction with a DNA molecule containing two sites in cis. The tetrameric enzyme is composed of four endonuclease subunits, each of which is symbolized by a rectangle. Two monomeric subunits form a pseudo-dimer and two of these pseudo-dimers are packed back-to-back forming two identical DNA binding clefts. With DNA containing two recognition sites, the enzyme forms intermediate DNA loops. The DNA is concertedly cut at both DNA recognition sites. (C) One of the type IIS REase interactions with DNA as exemplified by the best understood enzyme FokI.[87,88] The

circle represents the DNA recognition domain; while the rectangle represents the cleavage domain of the enzyme. The two enzyme monomers are yellow and white respectively. The enzyme dimerizes probably after each monomer through its DNA recognition domain has specifically bound to a DNA recognition site, because specific DNA binding results in opening of the FokI monomer.[87] This opening is most likely required for dimerization, which is mediated only by the cleavage domains. Two different complexes of a FokI dimer and two DNA recognition sites have been modeled depending on the relative orientation of the asymmetric recognition sites. The cleavage domains of two FokI monomers cooperate to cleave at either one of the two coordinated recognition sites that is in contact with both catalytic domains (9/13) bp downstream of this site (the first numeral in the parentheses indicates the position of cleavage in the strand written and the second numeral the cleavage position on the complementary strand).

1.3.3.2 Similarity in three-dimensional structures

Most type II restriction endonucleases, whose crystal structures have been determined (as shown in Figure 1.16), have a PD…D/EXK motif in the active site in common.[89,90] These restriction enzymes can be further divided in two groups: EcoRI-like family and EcoRV-like family as shown in Figure 1.17. Table 1.5 lists the available crystal structure information of 16 type II restriction endonucleases of the PD…D/ExK superfamily and seven other nucleases belonging to this family.

The structural similarity across the type II restriction endonucleases suggests that they have a common ancestor. The distinction between an EcoRI-like family and EcoRV-like family had been made not only from the associated similarities in structure but also with similarities in function: EcoRI, like BamHI, binds the DNA from the major groove side and produces sticky ends with 5′-overhangs, whereas EcoRV, like PvuII, approaches the DNA from the minor groove side and produces blunt ends. This has consequences for the positioning of the two active sites and, therefore, for the arrangement of the two subunits in the homodimer. Thus, the nature of the cleavage pattern rather than the DNA sequence recognized, appears to be the most important constraint on the mode of dimerization of restriction endonucleases.[89]

Figure 1.16A shows nine structures of free restriction endonucleases, and Figure 1.16B shows 13 structures of specific enzyme-DNA complexes. A comparison of these crystal and co-crystal structures illustrates that these enzymes possess a common endonuclease folding core harbouring the active site, which serves as an important structure stabilization factor ('stabilization centre').[91] This core comprises a five-stranded mixed β-sheet flanked by two α-helices.[92] These enzymes include orthodox type II restriction enzymes producing sticky ends with a 5′-overhang (BamHI, BglII, EcoRI, MunI, BsoBI), sticky ends with a 3′-overhang (BglI) or blunt ends (EcoRV, PvuII), as well as members of the type IIS (FokI), type IIE (NaeI) and type IIF (Cfr10I, NgoMIV) subdivisions as shown in Figure 1.16. The conserved core harbors the catalytic centre: it brings into spatial proximity two carboxylates, typically one aspartate and one glutamate, or one aspartate residue and one lysine residue.

For type II restriction endonucleases, only four β-strands are absolutely conserved within the common core. Two of these strands (β_2 and β_3 in EcoRI; β_d and β_e in EcoRV) contain the amino acid residues directly involved in catalysis. The second and third strand of the β-sheet serve as a scaffold for the catalytic residues of the PD…D/ExK motif. The other ones may be critical for formation of the β-sheet and the hydrophobic core. The fifth β-strand can be parallel (as in EcoRI) or antiparallel (as in EcoRV) to the fourth strand.

Enzyme	Subtype	Recognition site	Catalytic residues	PDB code (reference)
EcoRI-like				
BamHI	IIP	G$^{\downarrow}$CTAGG	Asp94;Glu111;Glu113	1BAM (apo) [146], 1BHM (+ spec. DNA) [211], 2BAM (+ spec. DNA, Ca^{2+}), 3BAM (product, Mn^{2+}) [154], 1ESG (+ non-spec. DNA) [212]
BglII	IIP	A$^{\downarrow}$GATCT	Asp84;Glu93;Gln95	1ES8 (apo) [213], 1DFM (+ spec. DNA, Ca^{2+}), 1D2I (+ spec. DNA, Mg^{2+}) [147]
BsoBI	IIP	C$^{\downarrow}$PyCGPuG	Asp212;Glu240;Lys242	1DC1 (+ spec. DNA) [134]
Bse634I	IIF	Pu$^{\downarrow}$CCGGPy	Asp146;Glu212;Lys198	1KNV (apo) [192]
Cfr10I	IIF	Pu$^{\downarrow}$CCGGPy	Asp134;Glu204;Lys190	1CFR (apo) [45]
EcoRI	IIP	G$^{\downarrow}$AATTC	Asp91;Glu111;Lys113	1ERI (+ spec. DNA) [214], 1QC9 (apo), 1QPS (product, Mn^{2+}), 1CL8 (+ mod. DNA), 1QRH (R145K) (+ spec. DNA), 1QRI (E144D) (+ spec. DNA)
EcoRII	IIE	$^{\downarrow}$CCWGG	Asp299;Glu337;Lys324	1NA6 [36]
FokI	IIS	GGATG$N_9^{\downarrow}$NNNN$^{\uparrow}$	Asp450;Asp467;Lys469	2FOK (apo) [62], 1FOK (+ spec. DNA) [78]
MunI	IIP	C$^{\downarrow}$AATTG	Asp83;Glu98;Lys100	1D02 (D83A) (+ spec. DNA) [141]
NgoMIV	IIF	G$^{\downarrow}$CCGGC	Asp140;Glu201;Lys187	1FIU (product, Mg^{2+}) [48]
Related to EcoRI				
TnsA (Tn7transposase)			Glu63;Asp114;Lys132	1F1Z (Mg^{2+}) [215]
EcoRV-like				
BglI	IIP	GCCNNNN$^{\downarrow}$NGGC	Asp116;Asp142;Lys144	1DMU (+ spec. DNA, Ca^{2+}) [155]
EcoRV	IIP	GAT$^{\downarrow}$ATC	Asp74;Asp90;Lys92	1RVE (apo) [130], 1AZ3, 1AZ4 (apo) [216], 4RVE (+ spec. DNA) [130], 1B95 (+ spec. DNA) [217], 1AZ0 (+ spec. DNA, Ca^{2+}) [216], 1B94 (+ spec. DNA, Ca^{2+}) [217], 1EOO (+ spec. DNA) [169],1RVA (prod, Mg^{2+}) [168], 1BSS (T93A) (+ spec. DNA, Ca^{2+}) [219],1B96 (Q69E) (+ spec. DNA), 1B97 (Q69L) (+ spec. DNA) [217], 1SUZ (K92A) (+ spec. DNA, Mg^{2+}), 1SX8 (K92A) (+ DNA, Mn^{2+}), 1STX, 1SX5 (K38A) (product, Mn^{2+}), [159], 1BSU, 1BUA (+ mod. DNA) [218], 1RV5 (+ interrupted DNA) [220], 1EO3, 1EON (+ mod. DNA) [169], 2RVE (+ non-spec. DNA) [130],1RVB (+ non-spec. DNA, Mg^{2+}) [168]
HincII	IIP	GTPy$^{\downarrow}$PuAC	Asp114;Asp127;Lys129	1KC6 (+ spec. DNA) [117] 1TW8 (+ spec. DNA, Ca^{2+}) [157], 1HXV (prod, Mg^{2+}) [158], 1XHU (product Mn^{2+}) [158],
MspI	IIP	C$^{\downarrow}$CGG	Asp99;Asn117;Lys119	1SA3 (+ spec. DNA) [79]
NaeI	IIE	GCC$^{\downarrow}$GGC	Asp86;Asp95;Lys97	1EV7 (apo) [39], 1IAW (+ spec. DNA) [40]
PvuII	IIP	CAG$^{\downarrow}$CTG	Asp58;Glu68;Lys70	1PVU (apo) [221], 1K0Z (apo) (Pr^{3+}), 1H56 (apo) (Mg^{2+}) [222], 1PVI (+ spec. DNA) [223], 1EYU (+ spec. DNA, pH 4.6) [156], 1F0O (+ spec. DNA, Ca^{2+}, crosslinked, pH 7.5) [156], 2PVI (+ mod. DNA) [224], 3PVI (D34G) (+ spec. DNA [219], 1NI0 (Y94F)
Related to EcoRV				
λ-exonuclease		non-specific	Asp119;Glu129;Lys131	1AVQ [225]
RecB endonuclease		non-specific	Asp1067;Asp1080;Lys1082	1W36 [226]
S.solfataricus Hjc resolvase		structure specific	Asp42;Glu55;Lys57	1HH1 [227]
MutH		$^{\downarrow}$GATC	Asp70;Glu77;Lys79	1AZO, 2AZO [81]
T7 endonuclease I		structure specific	Asp55;Glu65;Lys67	1FZR (E65K) [228] 1M0I [229] 1M0D (Mn^{2+}) [229]
VSR endonuclease		C$^{\downarrow}$TWGG	Asp51	1VSR [82] 1CW0 [83] 1OGD [230]

Table 1.5—Crystal structures of type II restriction endonucleases of the PD…D/ExK family and related enzymes (reprinted from Pingoud *et al.*[51]).

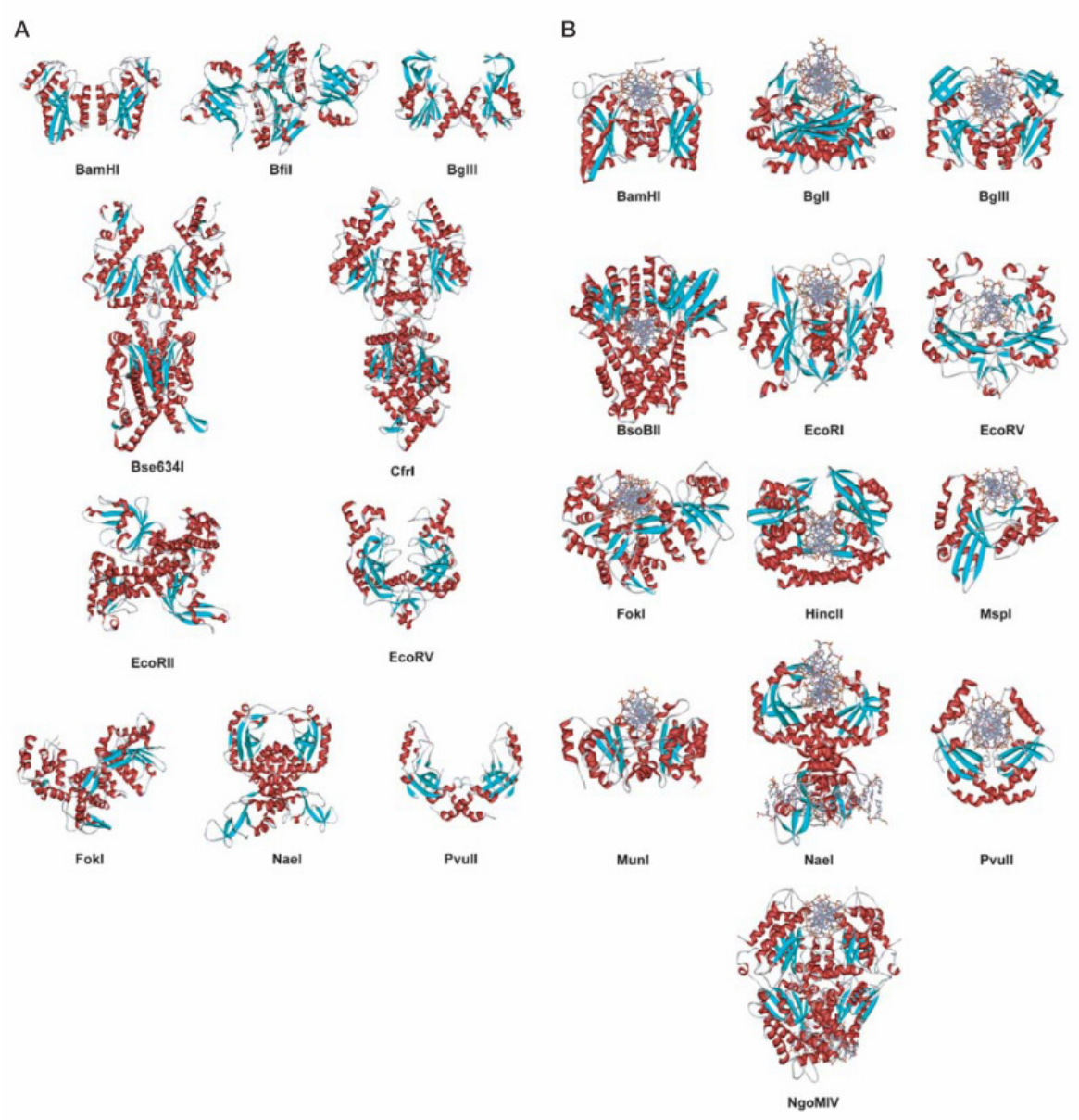


Figure 1.16—Crystal structures of some type II restriction endonucleases (reprinted from Pingoud *et al.*[51]). (A) free enzymes: BamHI (1BAM), BfiI (2C1L), BglII (1ES8), Bse634I (1KNV), Cfr10I (1CFR), EcoRII (1NA6), EcoRV (1RVE), FokI (2FOK), NaeI (11EV7), PvuII (1PVU); (B) specific restriction endonuclease-DNA complexes: BamHI (2BAM), BglI (1DMU), BglII (1DFM), BsoBI (1DC1), EcoRI (1ERI), EcoRV (4RVE), FokI (1FOK), HincII (1KC1), MspI (1SA3), MunI (1DO2), NaeI (1IAW), PvuII (1PVI), NgoMIV (1FIU). α-helices are depicted in red, β-strands are depicted in blue. Note that Bse634I, Cfr10I and NgoMIV are homotetrameric enzymes, EcoRII and NaeI as type IIE enzymes have an extra domain and FokI and MspI are monomeric enzymes in the co-crystal.

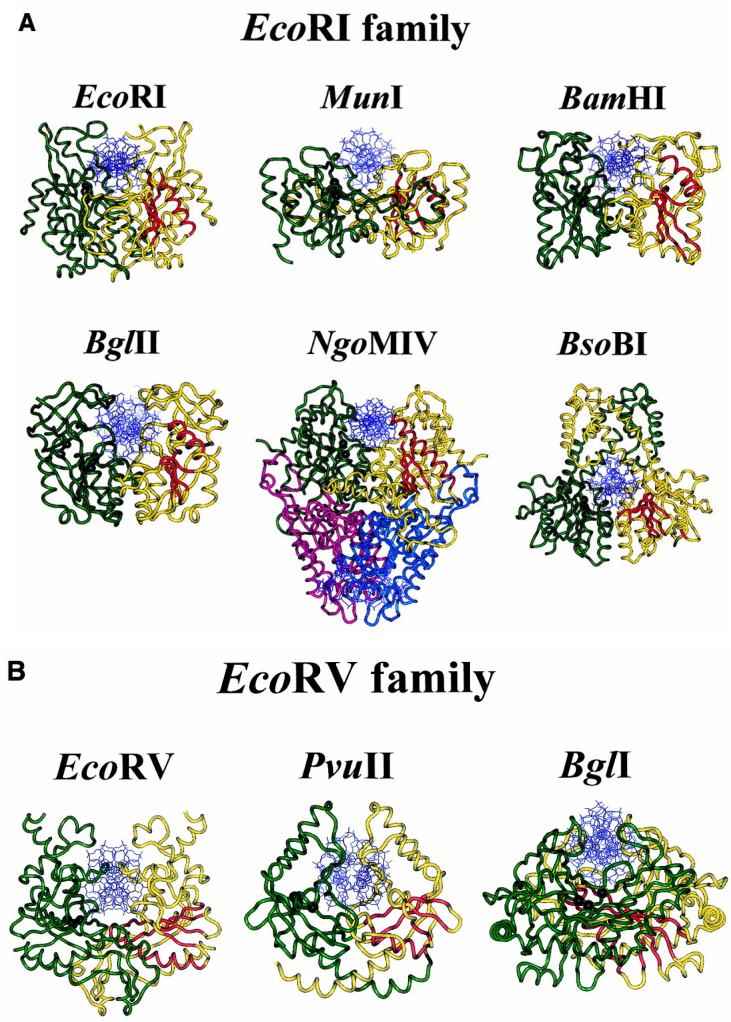


Figure 1.17—Crystal structures of specific type II restriction endonuclease-DNA complexes (reprinted from Pingoud *et al.*[65]). (A) EcoRI-like family. (B) EcoRI-like family. The two subunits of the homodimeric restriction endonucleases are shown in yellow and green (for the homotetrameric NgoMIV the individual subunits are colored yellow and green, purple and cyan), the DNA is shown in blue. In one subunit the four strictly conserved β-strands (EcoRI: β_1, β_2, β_3 and β_4; EcoRV: β_c, β_d, β_e and β_g) and one α-helix (EcoRI: α_2; EcoRV: α_B) of the common core are shown in red, in the other subunit the Cα-positions of the three essential amino acid residues of the PD...D/EXK motif are depicted as black spheres.

1.3.3.3 Steps for DNA binding and cleavage

Restriction endonucleases interact with DNA in a complex manner. Several specific steps need to be performed by orthodox type II restriction endonucleases during a catalytic cycle. Figure 1.18 presents a simple step-wise scheme involved in DNA

cleavage by a type II restriction endonuclease. The cycle starts with non-specific binding of the restriction enzyme to the macromolecular DNA, followed by a random diffusional translocation of the enzyme on the DNA molecule. When the enzyme reaches the recognition site of DNA, conformational changes happen as a part of the recognition process and lead to the activation of the catalytic centers. After the chemical reaction of phosphodiester bond cleavage in both strands, the product is released either by direct dissociation of the enzyme-product complex or by a transfer of the enzyme to non-specific sites on the same DNA molecule. Often this step is rate limiting for DNA cleavage by restriction enzymes under multiple turnover conditions.

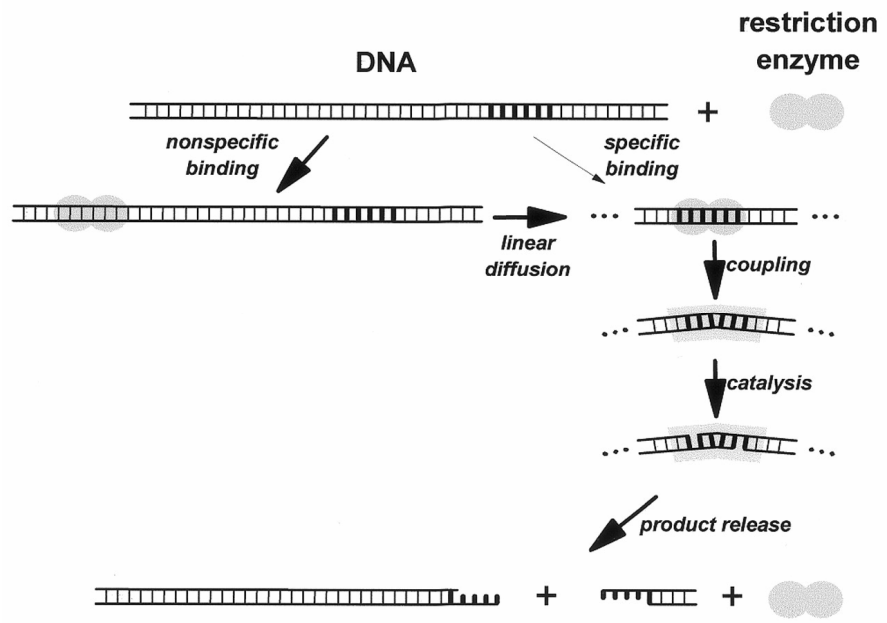


Figure 1.18—General reaction scheme for DNA binding and subsequent cleavage by type II restriction endonucleases (reprinted from Pingoud *et al.*[65]).

1.3.3.4 Target site searching

Before performing the cleavage activity on substrate DNA, all restriction enzymes encounter a problem: how they can efficiently find their respective recognition sequence in DNA molecules. By doing so, they have to effectively distinguish the recognition site from a large excess of non-specific sites.

A few principles of mechanism have been proposed to account for the efficiency of target site location by DNA-binding proteins in this regard: (i) 'sliding'; (ii) 'jumping' or 'hopping'; and (iii) intersegment transfer. It is important to note that these three principles are different, but not mutually exclusive.[93-95]

Sliding (also called linear or one-dimensional diffusion) describes the model where the protein stays bound to the DNA after the first encounter and moves along the DNA in a random manner, tracking the pitch of the double helix until it finds the specific site or dissociates.[96,97] It means that specific sites tend not to be missed by a restriction enzyme sliding along the DNA. During linear diffusion the non-specific binding mode is not given up and the water layer around DNA and protein, characteristic for the non-specific binding mode, remains largely intact.

Jumping or hopping is referred to as three-dimensional diffusion of protein, which is more inclined to re-associate with DNA molecules in a nearby region where it has dissociated rather than in a another DNA molecule or a distant site on the same molecule. Jumping and hopping does not follow the pitch of the double helix, meaning that specific sites along the DNA can be overlooked during hopping, depending on the step size. Small ligands binding to DNA, such as intercalating drugs, minor groove binders or triple-helix forming DNA, would not pose a major obstacle for jumping or hopping movements.

Intersegment transfer is only possible for proteins that have two DNA binding sites. If the DNA has one free binding site, the enzyme still remains bound to the DNA with the other binding site and is able to bind to the same DNA molecule at a distant location *via* its free DNA binding site. Binding of DNA on both DNA binding sites will generate loops in the DNA. Intersegment transfer is a particularly efficient way of scanning the DNA molecule over large distances. However, it is not a an effective means for covering small distances in search of a specific site due to steric constraints. Intersegment transfer is unlikely to be inhibited by small or even large DNA binding ligands.

Apart from the abovementioned three, a new mechanism with further modification has been proposed, with the application of single-molecule image techniques. Dikić *et al.*[98] used single-molecule images of EcoRV to study its mechanism and find its target sequence. They conjugated fluorescent labels of varying size (organic dyes, proteins and quantum dots) to EcoRV, and fused it to the engineered Rop protein scRM6. The linear diffusion constant (D1) was obtained by studying single-molecule images of these modified EcoRVs, revealing a significant size dependency. To account for the dependence of D1 on the size of the EcoRV label, they found that the experimental results can be best described by rotation-coupled sliding of the protein, which tracks the helical pitch of DNA as they diffuse along DNA.

1.3.3.5 Substrate recognition

For certain restriction enzymes, it is likely that the conformational changes will occur when binding to DNA non-specifically as evidenced by some crystal structure studies on EcoRV and BamHI shown in Figure 1.19. Obviously, for EcoRV, it opens the DNA-binding site, which requires a conformational change presumably triggered by a transient contact between the outer sides of the C-terminal arms of EcoRV and the DNA as reported by Schulze and Pingoud.[99] This conformational change enables DNA (non-specific or specific) to enter the binding cleft. Winkler and Wilson demonstrated that a region at the floor of the DNA binding site of EcoRV, which is disordered in the free enzyme, becomes ordered in the enzyme-DNA (non-specific) complex. The stable non-specific complex differs from the specific complex by being less compact and having a much smaller protein–DNA contact surface: 1370 versus 2173 Å.[100] Also it is notable that no base contacts are seen in the non-specific complex; DNA backbone contacts are fewer in number and differ substantially from those observed for the specific complex. In terms of BamHI, it seems as if major conformational transitions are not necessary for non-specific DNA to get access to the DNA binding cleft. Nevertheless a large segment at the floor of the DNA binding site encompassing residues 79-91 is dragged “down” by the nonspecific DNA as it drops out of the DNA binding cleft, indicating it is disordered in the free enzyme; while it becomes ordered in the non-specific complex. Also, this segment adopts a

different conformation here and is much more mobile than in the specific complex (C_α average B factor of 41 Å^2 versus 23 Å^2), reflecting far fewer electrostatic interactions with the DNA backbone. It is intriguing to note that no base-specific contacts and no direct DNA backbone contacts are seen in the non-specific complex, only a few water-mediated contacts even though the non-specific DNA used in the investigation only has one base pair variation from the specific DNA.[101] As observed for EcoRV, the non-specific BamHI-DNA complex is more open and less compact than the specific complex.

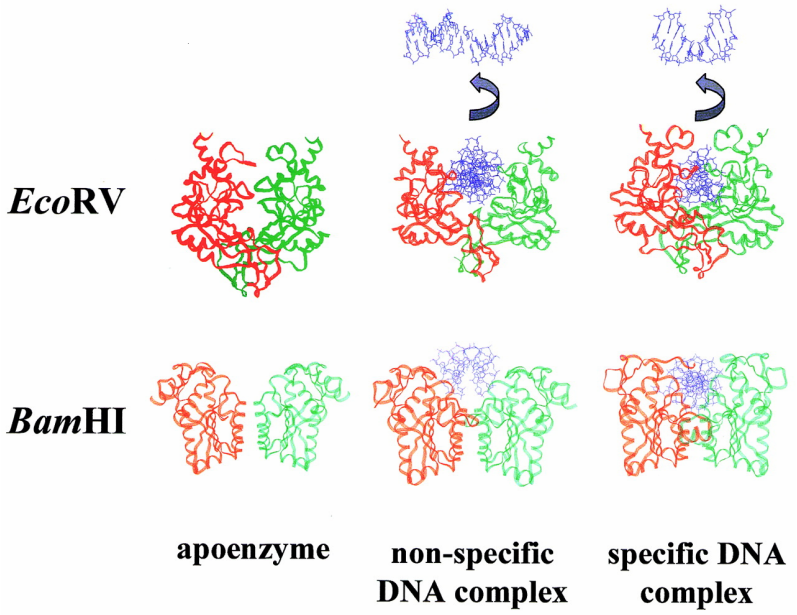


Figure 1.19—Comparison of the crystal structures of the free enzyme, the non-specific and the specific enzyme-DNA complex for the restriction endonucleases EcoRV and BamHI (reprinted from Pingoud *et al.*[65]). The two subunits are shown in red and green, the DNA in blue. For EcoRV, the side view of DNA in complex with enzyme is to demonstrate the absence of bending in the non-specific complex and the presence of bending in the specific complex. It is apparent that conformational changes take place when the free enzymes bind to non-specific DNA and to specific DNA, respectively. In the case of EcoRV, these conformational changes include the DNA which is bent by 50° in the specific complex. It is noteworthy that in the specific complex the DNA is almost completely encircled which gives the specific complex a more compact appearance compared to the non-specific one. PDB code numbers are: EcoRV (1RVE), non-specific complex (2RVE), specific complex (4RVE); BamHI (1BAM), non-specific complex (1ESG), specific complex (1BHM).

By examining the crystal and co-crystal structures available of specific restriction enzyme-DNA complexes, one can make some generalizations concerning the recognition process:[51,65]

(1) Most enzymes producing blunt ends (such as EcoRV) or sticky ends with 3'-overhangs (such as BglI) approach the DNA from the minor groove apart from HincII[102], whereas enzymes producing sticky ends with 5'-overhangs (such as EcoRI) contact the DNA from the major groove.

(2) Specific DNA-binding is accompanied by distortions of the DNA that bring functional groups of the enzyme into positions required for recognition. This also helps to position the scissile phosphates into the catalytic centre for supporting phosphodiester bond hydrolysis. Specific DNA binding is accompanied by protein conformational changes that involve structuring regions that are less ordered in the free enzyme or in the non-specific complex. These conformational changes often involve a repositioning of the subunits. In the specific complex the DNA is partially or as in most cases fully distorted. For example, in the cognate PvuII-DNA complex, the DNA duplex is not significantly bent and maintains a B-DNA-like conformation[103]. While for the specific HincII-DNA complex, protein-induced distortion of the DNA is accomplished by intercalation of glutamine side chains into the major groove on either side of the recognition site, which is the case for full DNA distortion in complex with enzyme.[102] BsoBI is an extreme case for fully distorting the cognate DNA upon enzyme binding. Each BsoBI monomer is formed by inserting a helical domain into an expanded EcoRI-type catalytic domain. DNA is completely encircled by a BsoBI dimer. Recognition sequence DNA lies within a 20 Å long tunnel of protein that excludes bulk solvent.[104]

(3) The formation of a highly co-operative hydrogen bond network is a characteristic feature for protein-cognate DNA complexes of restriction endonucleases. This hydrogen bond network comprises contacts to the bases ('direct read-out') as well as to the sugar-phosphate backbone ('indirect read-out') of DNA. A majority of the possible hydrogen bonds, mostly direct but also water-mediated, are formed to the

edges of the bases in the major groove and often in the minor groove. In addition, *van der Waals* interactions and hydrophobic contacts are formed to the bases of the recognition sequence. In general, several non-contiguous chain segments of a restriction enzyme are involved in direct and indirect readout. Whereas most of the specific contacts are between one subunit and one half-site of the palindromic recognition sequence, a few are directed to the other half-site.

1.3.3.6 Mechanism of phosphodiester bond hydrolysis

The chemical reaction for phosphodiester bond hydrolysis by type II restriction endonucleases adopts an nucleophilic substitution mechanism.[105,106] The general mechanism of phosphodiester hydrolysis comprises three steps:[51] (i) the preparation of the nucleophilic reagent by deprotonation,

$$H_2O + B + R'\text{-}O5'\text{-}(PO_2)^-\text{-}O3\text{-}R'' \longrightarrow OH^- + BH^+ + R'\text{-}O5'\text{-}(PO_2)^-\text{-}O3\text{-}R''$$

(ii) the nucleophilic attack of the hydroxide ion on the phosphorus resulting in formation of the pentavalent carbon center,

$$OH^- + BH^+ + R'\text{-}O5'\text{-}(PO_2)^-\text{-}O3\text{-}R'' \longrightarrow R'\text{-}O5'\text{-}(PO_2OH)^{2-}\text{-}O3\text{-}R'' + BH^+$$

(iii) the departure of the 3' hydroxyl leaving group

$$R'\text{-}O5'\text{-}(PO_2OH)^{2-}\text{-}O3\text{-}R'' + BH^+ \longrightarrow R'\text{-}O5'\text{-}(PO_2)\text{-}OH + HO3'\text{-}R'' + B$$

To achieve efficient catalysis, each step requires an assisting factor: (i) a base to deprotonate the water molecule; (ii) a Lewis acid that stabilizes the pentavalent transition state with two negative charges; and (iii) an acid that protonates the leaving 3'-oxyanion. The mechanistic models for DNA cleavage by restriction endonuclease are based on the number of metal ions involved in the reaction. Accordingly, three mechanisms with several variations were proposed: one-metal ion mechanism,[107] two-metal ion mechanism[108,109] and three-metal ion mechanism.[110]

1.3.3.7 Methyltransferase

Type II methyltransferases transfer a methyl group from SAM to a base in their respective DNA recognition site so as to protect DNA from being cleaved by their cognate restriction enzymes.[111] One of the most widely studied examples is M.HhaI from ***H****aemophilus* ***ha****emolyticus*. M.HhaI belongs to the class of m^5C-MTases. It binds to the four base sequence 5'-GCGC-3', and it methylates the cytosine residue towards the 5' end of this sequence. M.HhaI is one of the smallest (327 amino acids) and most studied of the m^5C-MTases. Kumar *et al.*[112] made the first observations of the crystal structure for M.HhaI in complex with SAM (AdoMet) cofactor. The structure was subsequently refined by Cheng *et al.*[113] Kilmasauskas *et al.*[114] firstly found that M.HhaI rotates the target methylated nucleotide out of the DNA helix into the HhaI catalytic pocket. This is commonly referred as 'base flipping', which has been inferred from various structural investigation obtained with different substrates and cofactors. (base flipping is reviewed in Section 1.4)

1.3.4 Type III R-M systems

Type III systems are similar to type I systems in a few ways. Extensively studied examples include for example EcoPI, EcoPI5, HinfIII. They are large, heteroligomeric complexes that can function as REases and MTases. They consist of two subunits: Mod subunit (~75 kDa), Res subunit (~106 kDa). The Mod subunit can act independently in recognising and modifying the DNA target sequence; it requires only SAM as a cofactor, and confers protection from restriction only by single strand methylation of the target recognition sequence. Methylation occurs at the N6 position of adenine. Res is only active for DNA restriction in a Res_2Mod_2 stoichiometry[115]. Restriction requires Mg^{2+} and ATP as cofactors and occurs when a type III enzyme interacts with two inversely oriented copies of its 5-6 bp asymmetric recognition sequence[116]. ATP is utilised for DNA recognition, DNA translocation and DNA restriction. The restriction reaction is stimulated by SAM[117], taking place at a fixed site on one side of the recognition sequence, 25-27 base pairs downstream.[53,59] DNA restriction is preceded by DNA translocation while the complex remains bound to its

recognition sequence.[60] Only after stalled DNA translocation and/or collision of two separate complexes does restriction occur.

Type III REases require the presence of two oppositely oriented DNA recognition sites.[116] Cleavage will occur only if both the sites are unmethylated, but is prevented even if one of the sites is modified. Hence, methylation on only one strand is sufficient in protecting the DNA from cleavage.[53,59,117] Unlike type I R-M enzymes, type III R-M enzymes dissociate from their recognition site after DNA cleavage, and can therefore turn over.[60]

The best studied examples are EcoP1I and EcoP15I.[59] With respect to EcoP15I, it has recently been shown that restriction occurs only after two enzymes have interacted with each other while still bound to their recognition sites, which may be separated by up to 3500 bp. Fast-scan atomic force microscopy (AFM) has shown that communication between the two enzymes occurs through diffusive DNA looping and ATPase driven translocation of the DNA between the enzymes.[118]

1.3.5 Type IV R-M systems

Type IV R-M systems, different from the other three R-M systems, only have REases but with no associated methyltransferase. These enzymes recognise and cleave DNA that has been modified (methylated, hydroxymethylated or glucosyl-hydroxymethylated).

DpnI was the first restriction enzyme discovered to require a methylated substrate from *Streptococcus pneumoniae*, which is able to cleave the sequence GATC only if the A is methylated. Other strains of *S. pneumoniae* produce DpnII which has the complementary specificity, recognising and cleaving the same sequence only if the sequence is not methylated. The DpnI and DpnII pair of enzymes (together with the methyltransferases associated with DpnII) present one of the clearest arguments in support of the idea that the main function of R-M systems is to protect cells from viral infection.[59]

Three systems have been found in *E.coli*: McrA, McrBC and Mrr. The mcrBC locus in *E.coli* K-12 lies adjacent to the hsd genes encoding for the type I R-M enzyme EcoKI, while the gene encoding the Mrr system lies on the other side of the hsd genes. Therefore, the genes for the three independent restriction systems are all clustered together in a small region of the *E.coli* genome, which is thus called the immigration control region. McrA is coded by the *e14* prophage-like element of many strains of *E.coli*.[59]

The best studied type IV system is McrBC (modified cytosine restriction) from *E. coli* K-12. McrBC requires Mg^{2+} as a cofactor and is the only nuclease known to use GTP as an energy resource for translocation and restriction of DNA.[119] This enzyme recognizes two copies of a dinucleotide sequence, comprised of a purine followed by a cytosine methylated at either the N4 or the C5 position, which are separated by between 40 to 3000 bp; and cleaves between methylated bases at multiple sites of DNA, preferentially 30 bp away from one of the sites.[120] As in type I R-M systems, McrBC remains bound to its recognition sites and reels in the DNA. The stalling of two translocating complexes gives rise to cleavage.

1.4 Base Flipping of R-M Systems

In R-M systems, base flipping is a very striking DNA-protein interaction feature, which became apparent with the first crystal structure of a DNA MTase in complex with substrate DNA[114]. These enzymes completely flip the target base out of the DNA helix and bind it into a hydrophobic pocket in the active site for the methyl group chemical transfer reaction as shown in Figure 1.20. Looking closer, it is clear that the target cytosine base for methylation in this enzyme-DNA-cofactor ternary complex is rotated, positioned perpendicularly to the long axis of the DNA duplex, by 180° out of the helix and into the protein active site, which is referred to as base flipping.[111,121,122] This was the first direct observation of 'base-flipping' by an enzyme. Figure 1.21 shows the structural distortion of the DNA helix caused by the base flipping by M.HhaI.

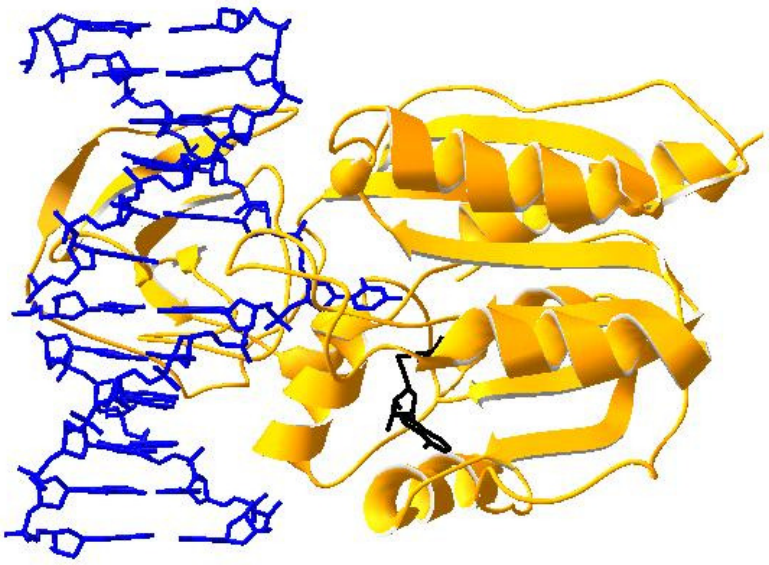


Figure 1.20—The X-ray crystal structure of the base-flipped DNA-M.HhaI-SAH ternary complex. The DNA is blue, the enzyme is orange and the SAH cofactor product is black (structure resolution is 2.60 Å, PDB: 1MHT).

Base flipping occurs through rotation of the phosphodiester bonds of the DNA backbone. The phosphorus-phosphorus bond distance is significantly larger in the flipped structure than in the stacked duplex. It is thought that this distortion allows the base to be rotated through the minor groove of the DNA and into the enzyme active site. In the DNA-M.HhaI-cofactor complex, there are no covalent bonds broken during base flipping, however the hydrogen bonding of the flipped base with its partner base and the stacking π-interaction with the adjacent base pairs is lost.

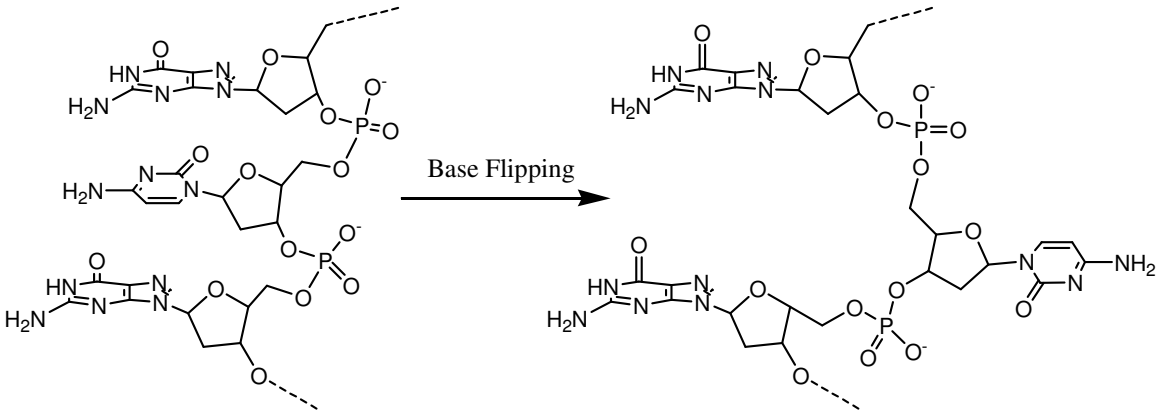


Figure 1.21—Structural distortion conferred on the DNA helix during base flipping (adapted from Klimasauskas *et al.*[123]).

Gara *et al.*[124] reported that in the HhaI Mtase in complex with an abasic site at the position replacing the target cytosine base, the enzyme still moves the sugar-backbone to the 'flipped-out' position. This phenomenon has also been validated in other DNA manipulating enzymes with their respective cognate sequence such as glycosylase, uracil DNA glycosylase (UDG), mismatch-specific uracil glycosylase (MUG), alkyladenine glycosylase (AAG) and alkylation glycosylase (AlkA). This indicates that the target base for these enzymes do not show anything special in terms of base flipping and thereby it has been concluded that it is the backbone that is targeted for flipping by the enzyme and the base is merely carried along with it.[122]

Structures have been determined for complexes between HhaI methyltransferase (M.HhaI) and oligonucleotides containing a G:A, G:U at the target base pair. The mismatched adenine, uracil sites are all flipped out of the DNA helix and located in the enzyme's active-site pocket, adopting the same conformation as in the flipped-out normal substrate. Thus rotation of the DNA backbone is probably the key to base flipping. In addition, some other bases rather than the natural base pairs located at the target position for HhaI also show a base flipping conformation as reported by Klimaauskas *et al.*[125] and Yang *et al.*[126] Therefore it is clear that the methylation reaction is sensitive to the base at the target position, but the base flipping step is not.

Klimasauskas *et al.*[123] used ^{19}F nuclear magnetic resonance (NMR) spectroscopy to investigate the dynamics of the base flipping mechanism. In their experiments, two cytosine bases in the duplex were labelled with the ^{19}F probe atoms; one was at the target site for base flipping and one was located away from the target base as a control. The study shows at least four distinct signals in the ^{19}F NMR spectra of the labelled duplex. These are attributed to the free DNA duplex, a binary DNA-HhaI complex incorporating a stacked target base, a DNA-HhaI complex with an extrahelical target base and the ternary DNA-HhaI-AdoHcy complex where the base is flipped and locked in the enzyme active site as shown in Figure 1.22.

$$HhaI+DNA \leftrightarrow HhaI\text{-}DNA_{stacked} \leftrightarrow HhaI\text{-}DNA_{flipped} \leftrightarrow HhaI\text{-}AdoHcy\text{-}DNA_{locked}$$

Figure 1.22—Suggested binding and conformational equilibria between HhaI, DNA and the AdoHcy cofactor analogue.

In the binary complex it is thought that the target cytosine exists in equilibrium between the flipped and unflipped conformations. In the DNA-HhaI complex, the target base is thought to occupy an ensemble of extrahelical positions that are simply stabilised by enzyme binding because the ^{19}F-NMR signal from the flipped base in the binary complex is distinct from that in the ternary complex. Addition of AdoHcy (SAH) to the system apparently 'locks' the target base into a well defined, extrahelical conformation in the enzyme active site. Simply put, for HhaI, base flipping comprises at least two steps: first, the base is rotated out of the DNA helix, but not tightly bound by the enzyme and then it can be locked and inserted into the hydrophobic binding pocket of the enzyme. Also, Allan and Reich[127] used total intensity stopped flow fluorescence measurements of DNA containing 2AP, allowing presteady-state real time observation of the base flipping transition. They found that following the rapid formation of an enzyme-DNA collision complex, a biphasic increase in total intensity was observed. The fast phase dominated the total intensity increase with a rate nearly identical to $k_{methylation}$, as determined by rapid chemical quench-flow techniques. These results provide the first direct and continuous observation of base flipping and show that at least two distinct transitions occur at the flipped base after complex formation.

In the structure of M.HaeIII-DNA, the substrate cytosine is flipped out from the DNA helix[128], as observed for M.HhaI. Moreover, the structure of a ternary complex between M.TaqI (an adenine-N6-specific DNA MTase), and its substrates (the specific DNA) and a non-reactive SAM analog, showed a flipped adenine[129]. However, the three available MTase-DNA complexes use three different approaches for stabilizing the DNA structure with a flipped base.[130]

M.HhaI inserts an amino acid side chain into the space left by the flipped cytosine to restore stacking interactions with the neighboring base pairs as well as hydrogen bonding to the `orphaned' guanine. M.HaeIII also inserts a side chain into the DNA

helix, but the side chain opens a gap in the DNA for the partner guanine re-pairing with a nearby cytosine. As a consequence, the outer guanine is left without a complementary base but forms one hydrogen bond with a sugar oxygen on the opposite strand. For M.TaqI, the amino acid corresponding to Gln 237 of M.HhaI and Ile 221 of M.HaeIII is Pro 393. Pro 393 does not protrude into the DNA helix, but makes a *van der Waals* contact to the methyl group of the orphaned thymine, which is shifted towards the center of the helix and partially fills the space left by the flipped adenine. Thus the DNA-stabilizing strategy generated by base flipping is fulfilled by varying combinations of inserted protein side chains (M.HhaI, M.HaeIII) and/or movement of the orphaned base (M.HaeIII, M.TaqI).

Apart from base flipping of MTase, recently, it has been found that certain kinds of restriction enzymes are capable of base flipping their specific substrate as described in Chapter 6 . The prevalence of this mechanism in R-M systems and some other DNA-binding enzymes demonstrates its importance.

The lack of crystal and co-crystal structures stimulates the attempts to understand phase identification of base flipping as described. An improved approach needs to encompass certain criteria. It needs to be non-destructive to the sample, to be non-disruptive to enzyme behaviour, to provide unambiguous results and to have a reasonable data collection time and be low cost and less labour-intensive etc. This discussion now turns to fluorescence-based methods by using 2AP since it potentially fulfils most of these requirements.

1.5 Aims of the Project

In this thesis, the author planned to develop and validate time-resolved fluorescence of 2AP as a practicable approach for studying the transient unzipping of matched DNA during the conformational dynamics process, which either exists as nucleic acid structure *per se* or is induced by nucleic acid protein manipulation. This method was then applied to study the conformational dynamics of 3WJ and to test the double-nucleotide unzipping mechanism of human flap endonuclease. Moreover, two

kinds of restriction enzymes were characterised and their base flipping mechanisms were tested. Additionally, their cutting features in terms of CAG repeats would be investigated, aiming to develop tools for exploring unusual DNA structures containing super-long CAG or CTG repeats, which it is proved to be highly linked to some neurodegenerative diseases.

1.6 References

1. J. D. Watson and F. H. Crick, *Nature*, 1953, **171**, 737-738.
2. R. E. Dickerson, H. R. Drew, B. N. Conner, R. M. Wing, A. V. Fratini and M. L. Kopka, *Science*, 1982, **216**, 475-485.
3. M. Ghorbani and F. Mohammad-Rafiee, *Nucleic Acids Res.*, 2011, **39**, 1220-1230.
4. R. E. Dickerson, *Nucleic Acids Res.*, 1989, **17**, 1797-1803.
5. R. Chandrasekaran and S. Arnott, *J Biomol Struct Dyn*, 1996, **13**, 1015-1027.
6. R. Chandrasekaran, A. Radha, H. S. Park and S. Arnott, *J. Biomol. Struct. Dyn.*, 1989, **6**, 1203-1215.
7. S. Arnott, R. Chandrasekaran, D. L. Birdsall, A. G. Leslie and R. L. Ratliff, *Nature*, 1980, **283**, 743-745.
8. S. Arnott, P. J. Bond and R. Chandrasekaran, *Nature*, 1980, **287**, 561-563.
9. A. Ghosh and M. Bansal, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2003, **D59**, 620-626.
10. D. C. Ward, E. Reich and L. Stryer, *J. Biol. Chem.*, 1969, **244**, 1228-1237.
11. F. Ernst, *J. Mol. Biol.*, 1959, **1**, 87-105.
12. L. C. Sowers, Y. Boulard and G. V. Fazakerley, *Biochemistry*, 2000, **39**, 7613-7620.
13. L. C. Sowers, G. V. Fazakerley, R. Eritja, B. E. Kaplan and M. F. Goodman, *Proc. Natl. Acad. Sci. U. S. A.*, 1986, **83**, 5434-5438.
14. A. Albert and H. Taguchi, *J. Chem. Soc., Perkin Trans.*, 1973.
15. A. Holmén, B. Nordén and B. Albinsson, *J. Am. Chem. Soc.*, 1997, **119**, 3114-3121.
16. S. O. Kelley and J. K. Barton, *Science*, 1999, **283**, 375-381.

17. M. A. O'Neill and J. K. Barton, *J. Am. Chem. Soc.*, 2004, **126**, 11471-11483.
18. C. R. Guest, R. A. Hochstrasser, L. C. Sowers and D. P. Millar, *Biochemistry*, 1991, **30**, 3271-3279.
19. J. T. Stivers, *Nucleic Acids Res.*, 1998, **26**, 3837-3844.
20. E. L. Rachofsky, E. Seibert, J. T. Stivers, R. Osman and J. B. A. Ross, *Biochemistry*, 2001, **40**, 957-967.
21. A. Haller, U. Rieder, M. Aigner, S. C. Blanchard and R. Micura, *Nat. Chem. Biol.*, 2011, **7**, 393-400.
22. T. Xia, *Curr. Opin. Chem. Biol.*, 2008, **12**, 604-611.
23. B. W. Allan and N. O. Reich, *Biochemistry*, 1996, **35**, 14757-14762.
24. H. Gowher and A. Jeltsch, *J. Mol. Biol.*, 2000, **303**, 93-110.
25. B. Holz, E. Weinhold, S. Klimasauskas and S. Serva, *Nucleic Acids Res.*, 1998, **26**, 1076-1083.
26. E. G. Malygin, A. A. Evdokimov, V. V. Zinoviev, L. G. Ovechkina, W. M. Lindstrom, N. O. Reich, S. L. Schlagman and S. Hattman, *Nucleic Acids Res.*, 2001, **29**, 2361-2369.
27. S. S. Szegedi, N. O. Reich and R. I. Gumport, *Nucleic Acids Res.*, 2000, **28**, 3962-3971.
28. S. Bheemanaik, J. M. Bujnicki, V. Nagaraja and D. N. Rao, *Biol. Chem.*, 2006, **387**, 515-523.
29. Y. V. R. Reddy and D. N. Rao, *J. Mol. Biol.*, 2000, **298**, 597-610.
30. T. J. Su, B. A. Connolly, C. Darlington, R. Mallin and D. T. F. Dryden, *Nucleic Acids Res.*, 2004, **32**, 2223-2230.
31. K. Liebert, A. Hermann, M. Schlickenrieder and A. Jeltsch, *J. Mol. Biol.*, 2004, **341**, 443-454.
32. L. B. Bloom, M. R. Otto, J. M. Beechem and M. F. Goodman, *Biochemistry*, 1993, **32**, 11247-11258.
33. R. A. Hochstrasser, T. E. Carver, L. C. Sowers and D. P. Millar, *Biochemistry*, 1994, **33**, 11971-11979.
34. C. M. Joyce, O. Potapova, A. M. DeLucia, X. Huang, V. P. Basu and N. D. F. Grindley, *Biochemistry*, 2008, **47**, 6103-6116.
35. A. Újvári and C. T. Martin, *Biochemistry*, 1996, **35**, 14574-14582.

36. S. S. Sastry and B. M. Ross, *Biochemistry*, 1996, **35**, 15715-15725.
37. J. J. Sullivan, K. P. Bjornson, L. C. Sowers and P. L. deHaseth, *Biochemistry*, 1997, **36**, 8005-8012.
38. P. O. Lycksell, A. Gräslund, F. Claesens, L. W. McLaughlin, U. Larsson and R. Rigler, *Nucleic Acids Res.*, 1987, **15**, 9011-9025.
39. T. M. Nordlund, S. Andersson, L. Nilsson, R. Rigler, A. Graeslund and L. W. McLaughlin, *Biochemistry*, 1989, **28**, 9095-9103.
40. M. A. Carpenter and A. S. Bhagwat, *Nucleic Acids Res.*, 2008, **36**, 5417-5425.
41. R. K. Neely, G. Tamulaitis, K. Chen, M. Kubala, V. Siksnys and A. C. Jones, *Nucleic Acids Res.*, 2009, **37**, 6859-6870.
42. R. H. Szczepanowski, M. A. Carpenter, H. Czapinska, M. Zaremba, G. Tamulaitis, V. Siksnys, A. S. Bhagwat and M. Bochtler, *Nucleic Acids Res.*, 2008, **36**, 6109-6117.
43. G. Tamulaitis, M. Zaremba, R. H. Szczepanowski, M. Bochtler and V. Siksnys, *Nucleic Acids Res.*, 2007, **35**, 4792-4799.
44. J. T. Stivers, K. W. Pankiewicz and K. A. Watanabe, *Biochemistry*, 1998, **38**, 952-963.
45. K. D. Raney, L. C. Sowers, D. P. Millar and S. J. Benkovic, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 6644-6648.
46. E. Y. M. Bonnist, *PhD thesis, The University of Edinburgh*, 2008.
47. R. K. Neely, D. Daujotyte, S. Grazulis, S. W. Magennis, D. T. F. Dryden, S. Klimasauskas and A. C. Jones, *Nucleic Acids Res.*, 2005, **33**, 6953-6960.
48. S. E. Luria and M. L. Human, *J Bacteriol*, 1952, **64**, 557-569.
49. G. Bertani and J. J. Weigle, *J Bacteriol*, 1953, **65**, 113-121.
50. W. Arber and D. Dussoix, *J. Mol. Biol.*, 1962, **5**, 18-36.
51. A. Pingoud, M. Fuxreiter, V. Pingoud and W. Wende, *Cell. Mol. Life Sci.*, 2005, **62**, 685-707.
52. M. R. Tock and D. T. F. Dryden, *Curr. Opin. Microbiol.*, 2005, **8**, 466-472.
53. G. G. Wilson and N. E. Murray, *Annu. Rev. Genet.*, 1991, **25**, 585-627.
54. A. Jeltsch, *ChemBioChem*, 2002, **3**, 274-293.
55. R. J. Roberts, T. Vincze, J. Posfai and D. Macelis, *Nucleic Acids Res.*, 2005, **33**, D230-D232.

56. R. J. Roberts, T. Vincze, J. Posfai and D. Macelis, *Nucleic Acids Res.*, 2010, **38**, D234-D236.
57. G. F. Vovis, K. Horiuchi and N. D. Zinder, *Proc. Natl. Acad. Sci. U. S. A.*, 1974, **71**, 3810-3813.
58. D. T. F. Dryden, *Nat. Struct. Mol. Biol.*, 2004, **11**, 804-806.
59. T. A. Bickle and D. H. Krüger, *Microbiol. Rev.*. 1993, **57**, 434-450.
60. D. T. F. Dryden, N. E. Murray and D. N. Rao, *Nucleic Acids Res.*, 2001, **29**, 3728-3741.
61. G. P. Davies, I. Martin, S. S. Sturrock, A. Cronshaw, N. E. Murray and D. T. F. Dryden, *J. Mol. Biol.*, 1999, **290**, 565-579.
62. D. T. F. Dryden, S. S. Sturrock and M. Winter, *Nat. Struct. Mol. Biol.*, 1995, **2**, 632-635.
63. C. K. Kennaway, J. E. Taylor, C. F. Song, W. Potrzebowski, W. Nicholson, J. H. White, A. Swiderska, A. Obarska-Kosinska, P. Callow, L. P. Cooper, G. A. Roberts, J.-B. Artero, J. M. Bujnicki, J. Trinick, G. G. Kneale and D. T. F. Dryden, *Genes Dev.*, 2012, **26**, 92-104.
64. S. Sistla and D. N. Rao, *Crit. Rev. Biochem. Mol. Biol.*, 2004, **39**, 1-19.
65. A. Pingoud and A. Jeltsch, *Nucleic Acids Res.*, 2001, **29**, 3705-3727.
66. W. Szybalski, S. C. Kim, N. Hasan and A. J. Podhajska, *Gene*, 1991, **100**, 13-26.
67. J. Bitinaite, D. A. Wah, A. K. Aggarwal and I. Schildkraut, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 10570-10575.
68. D. H. Krüger, D. Kupper, A. Meisel, M. Reuter and C. Schroeder, *FEMS Microbiol. Rev.*, 1995, **17**, 177-184.
69. C. C. Yang and M. D. Topal, *Biochemistry*, 1992, **31**, 9657-9664.
70. O. V. Piatrauskene, V. N. Tashlitskii, M. G. Brevnov, I. Bakman and E. S. Gromova, *Biokhimiia*, 1996, **61**, 1257-1269.
71. S. E. Halford, D. T. Bilcock, N. P. Stanford, S. A. Williams, S. E. Milsom, N. A. Gormley, M. A. Watson, A. J. Bath, M. L. Embleton, D. M. Gowers, L. E. Daniels, S. H. Parry and M. D. Szczelkun, *Biochem. Soc. Trans.*, 1999, **27**, 696-699.
72. L. M. Wentzell, T. J. Nobbs and S. E. Halford, *J. Mol. Biol.*, 1995, **248**,

581-595.

73. M. Deibert, S. Grazulis, G. Sasnauskas, V. Siksnys and R. Huber, *Nat. Struct. Biol.*, 2000, **7**, 792-799.

74. V. Siksnys, R. Skirgaila, G. Sasnauskas, C. Urbanke, D. Cherny, S. Grazulis and R. Huber, *J. Mol. Biol.*, 1999, **291**, 1105-1118.

75. S. Grazulis, M. Deibert, R. Rimseliene, R. Skirgaila, G. Sasnauskas, A. Lagunavicius, V. Repin, C. Urbanke, R. Huber and V. Siksnys, *Nucleic Acids Res.*, 2002, **30**, 876-885.

76. M. Mücke, D. H. Kruger and M. Reuter, *Nucleic Acids Res.*, 2003, **31**, 6079-6084.

77. K. Stankevicius, A. Lubys, A. Timinskas, D. Vaitkevicius and A. Janulaitis, *Nucleic Acids Res.*, 1998, **26**, 1084-1091.

78. P. C. Hsieh, J. P. Xiao, D. O'Loane and S. Y. Xu, *J Bacteriol*, 2000, **182**, 949-955.

79. H. Kong and C. L. Smith, *Nucleic Acids Res.*, 1997, **25**, 3687-3692.

80. H. Kong, *J. Mol. Biol.*, 1998, **279**, 823-832.

81. A. Janulaitis, R. Vaisvila, A. Timinskas, S. Klimasauskas and V. Butkus, *Nucleic Acids Res.*, 1992, **20**, 6051-6056.

82. S. Lacks and B. Greenberg, *J. Biol. Chem.*, 1975, **250**, 4060-4066.

83. L. S. Higgins, C. Besnier and H. Kong, *Nucleic Acids Res.*, 2001, **29**, 2492-2501.

84. R. D. Morgan, C. Calvet, M. Demeter, R. Agra and H. Kong, *Biol. Chem.*, 2000, **381**, 1123-1125.

85. Q. Song, H. Wu, F. Feng, G. Zhou, T. Kajiyama and H. Kambara, *Anal. Chem.*, 2010, **82**, 2074-2081.

86. P. Zhang, P. H.-M. Too, J. C. Samuelson, S.-H. Chan, T. Vincze, S. Doucette, S. Baeckstroem, K. D. Potamousis, T. M. Schramm, D. Forrest, D. C. Schwartz and S.-y. Xu, *Protein Expression Purif.*, 2010, **69**, 226-234.

87. E. S. Vanamee, S. Santagata and A. K. Aggarwal, *J. Mol. Biol.*, 2001, **309**, 69-78.

88. A. J. Bath, S. E. Milsom, N. A. Gormley and S. E. Halford, *J. Biol. Chem.*, 2002, **277**, 4024-4033.

89. J. E. Anderson, *Curr. Opin. Struct. Biol.*, 1993, **3**, 24-30.
90. V. Thielking, U. Selent, E. Koehler, H. Wolfes, U. Pieper, R. Geiger, C. Urbanke, F. K. Winkler and A. Pingoud, *Biochemistry*, 1991, **30**, 6416-6422.
91. M. Fuxreiter and I. Simon, *Proteins: Struct., Funct., Genet.*, 2002, **48**, 320-326.
92. C. Venclovas, A. Timinskas and V. Siksnys, *Proteins*, 1994, **20**, 279-282.
93. A. Graneli, C. C. Yeykal, R. B. Robertson and E. C. Greene, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 1221-1226.
94. D. M. Gowers, G. G. Wilson and S. E. Halford, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 15883-15888.
95. O. G. Berg, R. B. Winter and H. P. H. Von, *Biochemistry*, 1981, **20**, 6929-6948.
96. A. Jeltsch, J. Alves, H. Wolfes, G. Maass and A. Pingoud, *Biochemistry*, 1994, **33**, 10215-10219.
97. J. Sun, H. Viadiu, A. K. Aggarwal and H. Weinstein, *Biophys. J.*, 2003, **84**, 3317-3325.
98. J. Dikić, C. Menges, S. Clarke, M. Kokkinidis, A. Pingoud, W. Wende and P. Desbiolles, *Nucleic Acids Res.*, 2012.
99. C. Schulze, A. Jeltsch, I. Franke, C. Urbanke and A. Pingoud, *EMBO J.*, 1998, **17**, 6757-6766.
100. F. K. Winkler, D. W. Banner, C. Oefner, D. Tsernoglou, R. S. Brown, S. P. Heathman, R. K. Bryan, P. D. Martin, K. Petratos and K. S. Wilson, *EMBO J.*, 1993, **12**, 1781-1795.
101. H. Viadiu and A. K. Aggarwal, *Mol. Cell*, 2000, **5**, 889-895.
102. N. C. Horton, L. F. Dorner and J. J. Perona, *Nat. Struct. Biol.*, 2002, **9**, 42-47.
103. X. Cheng, K. Balendiran, I. Schildkraut and J. E. Anderson, *EMBO J.*, 1994, **13**, 3927-3935.
104. d. W. M. J. Van, J. J. Pelletier, S.-Y. Xu and A. M. Friedman, *Structure*, 2001, **9**, 133-144.
105. B. A. Connolly, F. Eckstein and A. Pingoud, *J. Biol. Chem.*, 1984, **259**, 10760-10763.
106. J. A. Grasby and B. A. Connolly, *Biochemistry*, 1992, **31**, 7855-7861.

107. A. Jeltsch, J. Alves, H. Wolfes, G. Maass and A. Pingoud, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 8499-8503.

108. L. S. Beese and T. A. Steitz, *EMBO J.*, 1991, **10**, 25-33.

109. M. Fothergill, M. F. Goodman, J. Petruska and A. Warshel, *J. Am. Chem. Soc.*, 1995, **117**, 11619-11627.

110. N. C. Horton and J. J. Perona, *Biochemistry*, 2004, **43**, 6841-6857.

111. X. Cheng and R. J. Roberts, *Nucleic Acids Res.*, 2001, **29**, 3784-3795.

112. S. Kumar, X. Cheng, J. W. Pflugrath and R. J. Roberts, *Biochemistry*, 1992, **31**, 8648-8653.

113. X. Cheng, S. Kumar, J. Posfai, J. W. Pflugrath and R. J. Roberts, *Cell*, 1993, **74**, 299-307.

114. S. Klimasauskas, S. Kumar, R. J. Roberts and X. Cheng, *Cell*, 1994, **76**, 357-369.

115. P. Janscak, U. Sandmeier, M. D. Szczelkun and T. A. Bickle, *J. Mol. Biol.*, 2001, **306**, 417-431.

116. A. Meisel, T. A. Bickle, D. H. Kriiger and C. Schroeder, *Nature*, 1992, **355**, 467-469.

117. A. A. Bourniquel and Thomas A. Bickle, *Biochimie*, 2002, **84**, 1047-1059.

118. N. Crampton, M. Yokokawa, D. T. F. Dryden, J. M. Edwardson, D. N. Rao, K. Takeyasu, S. H. Yoshimura and R. M. Henderson, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 12755-12760.

119. U. Pieper, T. Brinkmann, T. Krüger, M. Noyer-Weidner and A. Pingoud, *J. Mol. Biol.*, 1997, **272**, 190-199.

120. E. Sutherland, L. Coe and E. A. Raleigh, *J. Mol. Biol.*, 1992, **225**, 327-348.

121. D. R. Mernagh, I. A. Taylor and G. G. Kneale, *Biochem. J.*, 1998, **336**, 719-725.

122. R. J. Roberts and X. Cheng, *Annu. Rev. Biochem.*, 1998, **67**, 181-198.

123. S. Klimasauskas, T. Szyperski, S. Serva and K. Wuthrich, *EMBO J.*, 1998, **17**, 317-324.

124. M. O'Gara, J. R. Horton, R. J. Roberts and X. Cheng, *Nat. Struct. Biol.*, 1998, **5**, 872-877.

125. S. Klimasauskas and R. J. Roberts, *Nucleic Acids Res.*, 1995, **23**, 1388-1395.

126. A. S. Yang, J.-C. Shen, J.-M. Zingg, S. Mi and P. A. Jones, *Nucleic Acids Res.*, 1995, **23**, 1380-1387.

127. B. W. Allan, J. M. Beechem, W. M. Lindstrom and N. O. Reich, *J. Biol. Chem.*, 1998, **273**, 2368-2373.

128. K. M. Reinisch, L. Chen, G. L. Verdine and W. N. Lipscomb, *Cell*, 1995, **82**, 143-153.

129. K. Goedecke, M. Pignot, R. S. Goody, A. J. Scheidig and E. Weinhold, *Nat. Struct. Biol.*, 2001, **8**, 121-125.

130. R. M. Blumenthal and X. Cheng, *Nat. Struct. Biol.*, 2001, **8**, 101-103.

Chapter 2---Materials and Methods

2.1 Materials

2.1.1 Chemicals and oligonucleotides

Broad range pre-stained molecular mass markers for SDS-PAGE were purchased from BioRad (Precision Plus protein standards; Hercules, USA). DNA ladder was purchased from Promega (Madison, WI, USA). All other chemicals were purchased from Sigma-Aldrich (Dorset, UK), unless otherwise stated. *S*-adenosyl-L-methionine and *S*-adenosyl-L-homocysteine were purchased from Fluka Chemie GmbH (Buchs, Switzerland). The sequences of the oligonucleotides can be found in the 'Experimental' section in individual chapter. All solutions were made up in distilled, de-ionised water, except HPLC gel filtration buffer which used HPLC grade water. Regular oligonucleotides were purchased from Sigma-Aldrich custom oligonucleotides service. 5'-hexachlorofluorescein (HEX)- and black hole quencher 1 (BHQ1)-labelled oligonucleotides and some of the 2AP-labelled oligonucleotides were purchased from ATDbio Ltd (Southampton, UK). The other 2AP-labelled oligonucleotides were synthesised by the author.

2.1.2 Reagents

DNA loading buffer (10×): 20% Ficoll 400, 0.1 M Na_2EDTA (pH 8.0), 1% SDS, 0.25 % (w/v) bromophenol blue.

SDS-PAGE de-staining solution: 10% acetic acid; 30% methanol; 60% H_2O.

SDS-PAGE staining solution: 20% acetic acid, 50% methanol, 29.75% H_2O, 0.25 % (w/v) Coomassie Brilliant Blue.

SDS Laemmli sample buffer (2×): purchased from Sigma-Aldrich.

TBE (10×): 0.89 M Tris base; 0.89 M boric acid; 20 mM EDTA.

TE buffer: 10 mM Tris-HCl, 1 mM EDTA, pH 8.0.

TAE buffer: 40 mM Tris-HCl, 2 mM EDTA, 24 mM acetic acid, pH 7.7.

SDS sample buffer (5×): 10% (w/v) SDS, 10 mM β-mercaptoethanol, 20% (v/v) glycerol, 0.2 M Tris-HCl, pH 6.8, 0.05% (w/v) bromophenol blue.

2.2 Methods

2.2.1 DNA methods

2.2.1.1 DNA synthesis

Classically, the first base of DNA of the target sequence to be synthesized was incorporated in a based-functionalised controlled-pore glass (CPG) or polystyrene support. Chemical synthesis can be initialised by removing (deblocking or detritylation) the 5'-dimethoxytrityl (DMT) group with use of acid which is usually 3% trichloroacetic acid (TCA) in dichloromethane (DCM) to restore the reactive OH group. The phosphoramidite corresponding to the second base in the sequence was activated using tetrazole-like products such as ETT (5'-ethylthio-1H-tetrazole) or BTT (5-benzylthio-1H-tetrazole), which can then be coupled to the first base *via* a 5'-OH group by forming a phosphate linkage. Usually, the phosphoramidite coupling can reach around 99% efficiency. However, the 1% remaining reactant with reactive 5'-OH would result in unwanted side products if it is not properly treated. In a bid to avoid this, a capping procedure was introduced to acetylate the unreacted 5'-OH before further reaction takes place. This was done using acetic anhydride and the catalyst N-methylimidazole solution. After this, trivalent phosphate triester linkage was oxidised to the more stable pentavalent phosphotriester using iodine solution. One synthesis cycle was complete at this stage and more cycles needed to be repeated for the subsequent bases, starting with detritylation of the second molecule and so on. Several cycles needed to be repeated until the DNA required was finished. At this point, two alternative options were available: either leaving the DMT group on (DMT-on) or removing it (DMT-off) by a final acid treatment. The synthesised DNA was cleaved from the solid support (column) by soaking in ammonia solution at an elevated temperature (55-60°C) for several hours. This step also deprotected phosphorus *via* β-elimination of the cyanoethyl group and removed the protection groups on the bases. However, this cleavage and deprotection is likely to vary upon the different base protection and modifier used.

.

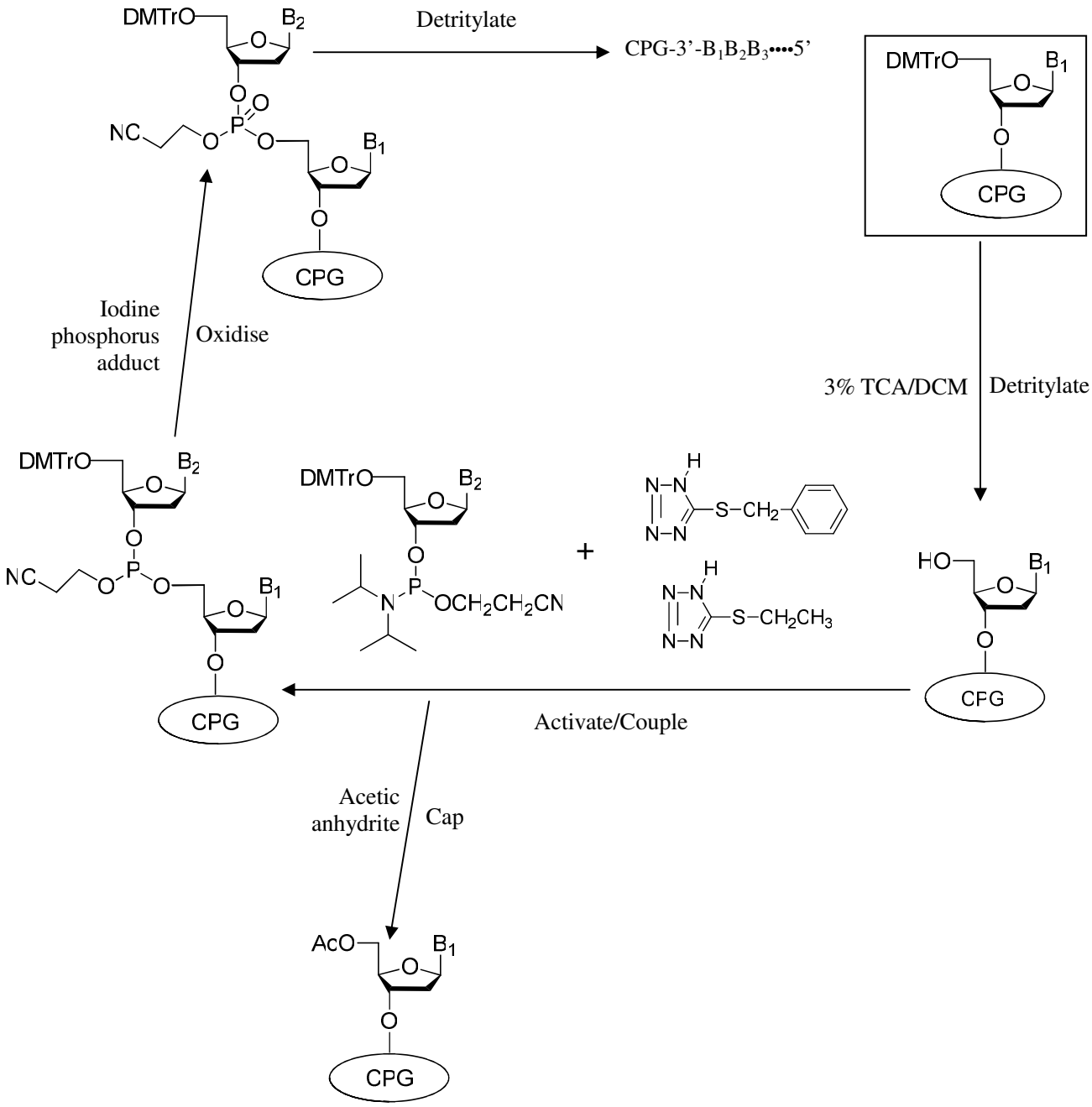


Figure 2.1—Scheme of oligonucleotides synthesis cycle using phosphoramidite chemistry.

2.2.1.2 Purification of chemically synthesised DNA

Purification of chemically synthesised DNA was done by high performance liquid chromatography (HPLC). Leaving the DMT group on the synthesised product renders a convenient 'handle' for purification, since the hydrophobic DMT group will elongate the retention time in reverse phase chromatography. Only the full-length product has the DMT group while the truncation doesn't; therefore a separation can be made between them.

Firstly, the synthesised DNA was cleaved from the solid column and the CPG was left to stand for 1 hour at room temperature to cleave the DNA from the support. This room-temperature step allowed enough time to cleave the DNA from the column whilst avoiding dissolving the CPG. The solid support was then filtered out through the spin vial using a centrifuge, collecting the solution containing the oligonuceotides. This DNA solution was further processed at 60°C for 2 hours to restore the biologically active DNA. This DNA solution was concentrated down to a volume of 400 μL using a Speedvac, and was then ready for HPLC purification.

A standard two-stage DMT (dimethoxytrityl)-on/DMT-off reverse phase HPLC purification was used. Firstly, an analytical amount of sample was injected and run on the HPLC and an HPLC trace containing two main peaks was obtained. The first peak corresponded to the truncation and the second peak the full-length product. A similar but preparative scale HPLC injection was made after this and the fractions of the second peak collected. The collected sample was then deprotected by leaving it with an equal amount of 40% acetic acid/water (resulting in a 20% (v/v) final concentration of acid) at room temperature for 1.5 hours. After waiting 1.5 hours for the DMT-groups to be deprotected, the sample was then transferred to a Speedvac and concentrated down to a volume of less than 0.5 mL.

The sample was then injected onto the HPLC using a DMT-off protocol, where the purity of the DNA could be seen. The collected sample after HPLC was further desalted using NAP-10 columns. The Speedvac was used to concentrate the

oligonucleotides eluted down to ~100 µL and the concentration of the sample was measured using a UV-Vis spectrometer.

2.2.1.3 Characterization of synthesised DNA by mass spectrometry (MS)

Characterization of synthesised DNA was done by electrospray ionization fourier transform ion cyclotron resonance (ESI-FTICR) MS. Mass analysis was acquired using the Bruker solariX Fourier transform ion cyclotron resonance mass spectrometer (Bruker Daltonics, Billerica, MA, USA) equipped with 12.0 T superconducting magnet. Sample introduction was achieved using the Advion TriVersa NanoMate operating in negative ion mode, facilitating simultaneous nano-electrospray ionization (nESI) and fraction collection. The infusion solution used was 7 mM triethylammonium acetate (TEAA), 14% acetonitrile, 14% 2-propanol and 6% MeOH. The molecular weights (M.W.) of the oligonucleotides were derived from the deconvoluted MS using the maximum entropy method integrated in the analysis software. Monoisotopic, average and nominal molecular mass were all used for comparison with the M.W. from the MS.

2.2.1.4 Determination of the DNA extinction coefficient

The method adopted for the DNA extinction coefficient calculation was based on the nearest-neighbor model and its published parameters.[1] These parameters are determined at a wavelength of 260 nm, 25°C, and neutral pH. The average error of calculated extinction coefficients was shown to be within 4% under these conditions.[2,3]

The estimated extinction coefficient would be accurate in a pH range from 6.5 to 8.5. When measuring DNA concentration, it is therefore recommended to use a low salt buffer to maintain neutral pH. Metal cations (Cu^{2+}, Hg^{2+}) that bind to bases can dramatically change the DNA extinction coefficient.[4] Magnesium cations associate mostly with negatively charged phosphate backbone groups. Mg^{2+} ions are known to slightly decrease the DNA extinction coefficient (5-10%). Integrated DNA

Technologies biophysics online software[5] (http://biophysics.idtdna.com/) was used to predict the UV spectra and extinction coefficients for both ssDNA and dsDNA. The concentration of oligonucleotide thus can be determined using extinction coefficients calculated based on the Lambert-Beer law.

2.2.2 Protein methods

2.2.2.1 Determination of the extinction coefficient of protein

The molar extinction coefficient (ε) is the parameter representing how much light a protein absorbs at a certain wavelength per molar concentration. It is useful to have an estimation of this coefficient for following a protein with a spectrophotometer when purifying it. The units for the molar extinction coefficient can be described as $M^{-1}cm^{-1}$. The molar extinction coefficient is equal to the absorption of a one molar sample in a 1 cm light path.

The molar extinction coefficient of a protein is constant at a specific wavelength. In biochemistry, the extinction coefficient of a protein at 280 nm is commonly used. It has been reported that one can estimate the molar extinction coefficient of a protein from its amino acid composition. Different amino acids have a different absorption at 280 nm. Tryptophan and tyrosine have a large absorption at 280 nm wavelength. Cysteine does not absorb substantially at wavelengths >260 nm, while cystine, (the oxidation state where two cysteine residues are covalently linked by a disulfide bond) does. The theoretical extinction coefficient of a given protein can be calculated using the number of tryptophan, tyrosine and cysteine amino acid residues as reported previously.[7,8] The extinction coefficient of the native protein in water can be calculated using the following equation:

Ext(Prot)= Numb(Tyr)×Ext(Tyr)+Numb(Trp)×Ext(Trp)+Numb(Cystine)×Ext(Cystine)

Where (for proteins in water measured at 280 nm):

Ext(Tyr)=1490 $M^{-1}cm^{-1}$

Ext(Trp)=5500 $M^{-1}cm^{-1}$

Ext(Cystine)=125 $M^{-1}cm^{-1}$

The absorbance (optical density) can be calculated using the following formula:

Absorbance(Prot)=Ext(Prot)/Molecular weight

Note: Ext=Extinction coefficient; Numb=number; Tyr=tyrosine; Trp=tryptophan; Prot=protein

2.2.2.2 SDS polyacrylamide gel electrophoresis

SDS (Sodium dodecyl sulfate) polyacrylamide gel electrophoresis (PAGE) generally separates denatured proteins based on their molecular size. The protein samples were prepared by mixing an equal volume of 5×sample buffer. NuPAGE 4-12% Bis-Tris gels (Invitrogen, UK) were used. The running buffer used was NuPAGE MES SDS running buffer (Invitrogen, UK). The voltage and time employed for running the sample varied (see in experimental sections in different chapters). The SDS-PAGE gels were removed from their casts and gently soaked in staining solution for at least 30 min. The gels were destained with de-staining solution in a box. Then the box was put in microwave oven for 30 seconds in order to heat the gel solution and speed up the destaining process. Gels were destained for 3-5 hours or longer until bands can be clearly distinguished, replenishing the solution every hour. Images of the gels were captured using a digital camera (Fujifilm FinePix 602Zoom) and optimised using Photoshop.

2.2.3 High-performance liquid chromatography (HPLC) methods

2.2.3.1 Analytical denaturing HPLC

A Gilson HPLC equipped with absorption detection at 254 nm was used for analysis of DNA cleavage as described before[9]. The reverse phase column (C18 Jones Chromatography) was kept at a constant 65°C. A linear acetonitrile gradient from 5% to 65% acetonitrile was generated by mixing of 0.1 M acetic acid, 5% acetonitrile and 0.1 M acetic acid, 65% acetonitrile aqueous solutions. The pH of the two solvents was adjusted to 6.5 with triethylamine.

2.2.3.2 Size-exclusion HPLC

Size-exclusion HPLC (SE-HPLC or SEC), also referred to as gel filtration, separates analytes based on a combination of their hydrodynamic size, diffusion coefficient, and surface properties. The stationary phase of SEC consists of inert particles packed into a dense three-dimensional matrix within a glass or steel column. The mobile phase can be purified water, an aqueous buffer, an organic solvent, or a mixture of these. The stationary-phase matrix has a certain dimension of pores, which only allows species below a certain size to enter. Large proteins in size are excluded from these pores, and their limited interaction with the stationary phase results in a short elution path. The smaller proteins, which can freely go into the pores, are removed from the flowing mobile phase, delaying their passage through the column. The time they spend in the stationary-phase pores is extended and their removal from the mobile phase flow causes them to take longer to elute and results in a separation between the particles based on differences in their size.

A Biosep-sec-s 3000 gel filtration column (Phenomenex, USA) was firstly calibrated with standard proteins and the calibration curve was plotted for subsequent use. Apoferritin (443 kDa), β-amylase (200 kDa), alcohol dehydrogenase (150 kDa), bovine serum albumen (66 kDa) and carbonic anhydrase (29 kDa) were individually employed as molecular mass markers for calibrating the SEC chromatographic column. All of the standard proteins are known to be globular. Details for experiments running different protein samples can be found in the 'Experimental' section of the individual chapter concerned.

2.2.4 Fluorescence methods

2.2.4.1 Fluorescence anisotropy

Steady-state fluorescence anisotropy measurement is valuable for monitoring the rotational dynamics of a macro-biomolecule in solution. Analysis of acquired data would be informative in terms of the macromolecule's size, shape and possible interactions with other macromolecules. Fluorescence anisotropy is reasonably

sensitive; and more importantly, it is a solution-based technique thus offering a versatile method to conduct equilibrium measurement of a binding in physiological solution condition. This makes it ideal for the study of protein–nucleic acid or protein–protein interactions.

In isotropic solution, all the fluorophores are oriented randomly. When such solution is irradiated with plane-polarised light, only certain fluorophores are excited. The selectively excited fluorophores are those having a component of the absorption transition dipole parallel to the electric vector of the excitation light. This selective excitation gives rise to a partially oriented population of fluorophores and also partially polarized fluorescence emission.

As shown in Figure 2.2, when irradiating the sample with vertically polarised light, whose electric vector is oriented parallel to the vertical or z-axis, by measuring the intensity of the emission through a polarizer, the anisotropy can be obtained. When the emission polarizer is set parallel to the direction of the polarised excitation, the intensity measured is termed as $I_{\parallel}$, while when the emission polarizer is perpendicular to the excitation, the intensity is called $I_{\perp}$. The equation used for calculating the anisotropy, r, is as follows:

Equation 2.1

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2 I_{\perp}}$$

The denominator describes total intensity of emission.

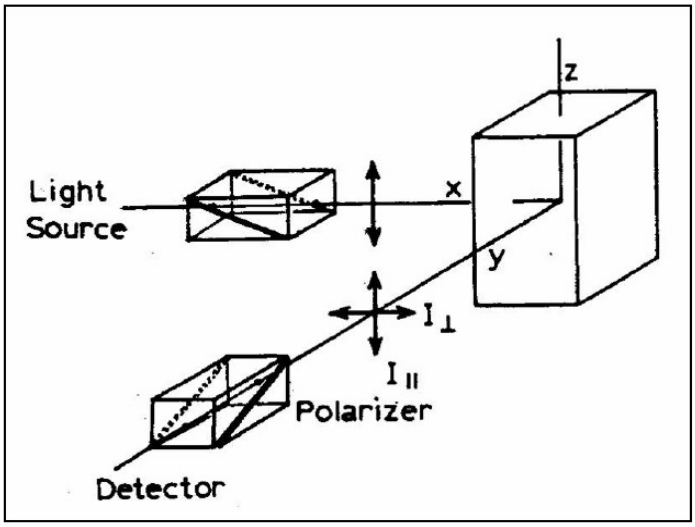


Figure 2.2—Schematic diagram showing fluorescence anisotropy measurement (adapted from Lakowicz[10]).

Anisotropy is a unit-less quantity that can be interpreted in terms of the extent to which a fluorophore rotates within the excited-state lifetime. Larger biomolecules will rotate more slowly due to greater mass and volume[11]; thus, if a biomolecule binds to another biomolecule, the rotational rate ('tumbling') will decline, giving rise to an increased anisotropy value[10].

Monochromators in fluorescence spectrometers can diffuse polychromatic light into its various wavelengths. The transmission efficiency of a monochromator is related to the polarisation displayed by the gratings of the monochromator, which means that the polarisation of the gratings can display different efficiencies of transmission among different wavelengths. This can have important consequences in the measurements of fluorescence anisotropy when the vertically and horizontally polarised emission is measured simultaneously. These inaccuracies caused by the varying efficiencies of these monochromators can be corrected by calculating the *G* factor.

The most two commonly used approach to measure anisotropy are L-format and T-format setting-up. In the L-format, a single emission channel is used; whereas in the T-format, the parallel (*VV*) and the perpendicular (*VH*) components of the

emission spectra are measured simultaneously through separate channels. In the T-format measurements of fluorescence anisotropy, the intensities of the parallel and perpendicular components of the emission spectrum are measured simultaneously using two separate detection systems (Figure 2.3). The *G* factor can be easily measured by keeping the excitation beam horizontally polarised (excitation angle= 90°), and measuring the emission spectra of the vertically polarised emission (I_{HV}, emission angle=0°) as well as the horizontally polarised emission (I_{HH}, emission angle = 90°) simultaneously. The ratio of the two intensities gives the *G* factor.

Equation 2.2

$$G = \frac{I_{HV}}{I_{HH}}$$

where I_{HV}=intensity of vertically polarised emission spectra measured (0°) when excited with a horizontally polarised (90°) beam; I_{HH}=intensity of horizontally polarised emission spectra measured (90°) when excited with a horizontally polarised (90°) beam.

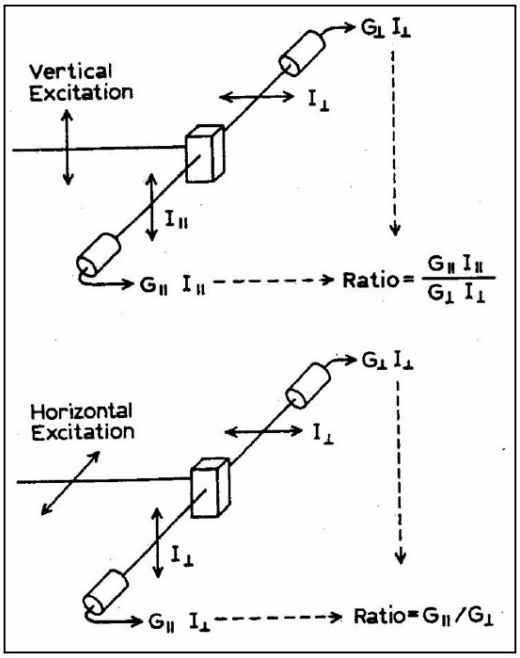


Figure 2.3—Schematic arrangements for T-format measurements of fluorescence anisotropy. $G_{\parallel}$ and $G_{\perp}$ are the gains of the parallel and perpendicular channels. (adapted from Lakowicz[10]).

The *G* factor of the system is recorded at the beginning and at the end of every single experiment and is adjusted to a value of roughly 1.0. When the *G* factor is known, the anisotropy (r) of the sample can be calculated by:

Equation 2.3

$$r = \frac{I_{VV} - GI_{VH}}{I_{VV} + 2GI_{VH}}$$

where **G**=$I_{HV}/I_{HH;}$ I_{VV}=fluorescence intensity of the vertically polarised emission (*V* =0°), when sample is excited with vertically polarised light (*V*=0°); I_{VH}= fluorescence intensity of the horizontally polarised emission (*H*=90°), when the sample is excited with vertically polarised light (*V*=0°).

As for conducting the experiments, a 1 cm path length quartz cuvette (1×0.4 cm; Starna) containing ~800 µL of sample solution was placed in the fluorimeter and left for a few minutes, allowing the temperature equilibration. The magnetic stirrer was set to ~400 rpm to induce gentle mixing of the sample solution. Buffer used was filtered for purity and degassed beforehand for avoiding bubbles. The water bath of the fluorimeter, which circulates water through the cuvette holder was set at 25ºC. Excitation and emission slits were set at 10 nm. The measurement was performed using a T-format geometry.

Step 1:

The *G* factor was recorded by exciting (E_x=535 nm) with horizontally polarised light (90º) and emitting (E_m=535 nm) for both arms. It was found that the *G* factor was approximately equal to 1.

Step 2:

The following emission scans were recorded to assess the hexachlorofluorescein (HEX) fluorophores (chemical structure is shown in Chapter 6). I_{VV}, I_{VH}, I_{HH}, I_{HV}; E_x= 535 nm, E_m=555 nm, step=1 nm, dwell time=0.1 second, 2 scans was performed.

Step 3:

The *G* factor was recorded by a 'kinetic scan' model; I_{HH}, I_{HV} for 60 seconds; E_x= 535 nm, E_m=555 nm.

Step 4:

The readings were taken as the TseI enzyme was titrated into the sample solution containing the DNA duplex of interest, using a Hamilton syringe. The time for 'kinetic scan' was set for ~40 mins long enough for titration completion; I_{VV}, I_{VH}; E_x=535 nm, E_m=555 nm. At each titration point the lid of the sample chamber was removed and a protein aliquot was added into and quickly mixed in solution. Step 5 The *G* factor was calculated by repeating Step 3 after the titration was done.

2.2.4.2 Steady-state fluorescence measurements

Steady-state fluorescence measurements were conducted on a Horiba Jobin Yvon FluoroMax fluorimeter, equipped with a 150 W xenon lamp and separate excitation and emission monochromators or Edinburgh Instrument FS900 photon counting spectrofluorometer. To measure an emission spectrum, the excitation monochromator was preset at a single wavelength. The emission was scanned across an appropriate wavelength range. The emission spectrum shows the emission signal intensity against different emission wavelengths. The slit width was set at 5-10 nm, as appropriate, for both monochromators corresponding to the wavelength resolution for the chosen excitation and emission wavelengths. Integration time (dwell time), number of scans, scan step (in nanometers), were varied to obtain a reasonably smooth spectral trace. Finally, measured signal intensities were internally corrected for wavelength-dependent lamp output and for background by the Instruments SA, Inc. DataMax software. Samples in the aqueous solution were measured in a Starna fused silica fluorescence cuvette with a capacity ranging from 100 μL to 1000 μL as described in each chapter. Details for each experiment can be found in the individual chapter concerned.

2.2.4.3 Time-resolved fluorescence measurements

Fluorescence lifetime measurements were all carried out using the time-correlated single photon counting technique (TCSPC), in the Collaborative Optical Spectroscopy Micromanipulation and Imaging Centre (COSMIC), University of Edinburgh. A schematic diagram of the experimental setup is shown in Figure 2.4.

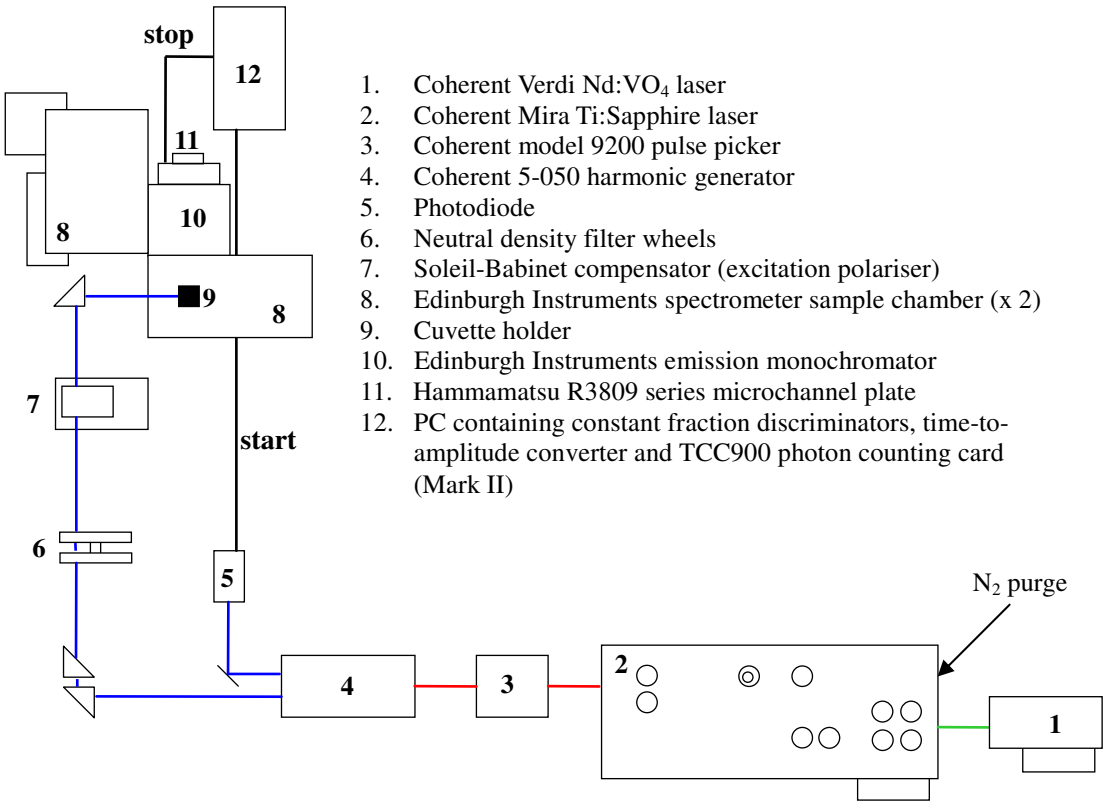


Figure 2.4—Schematic diagram of equipment used for time-resolved fluorescence experiments (adapted from Bonnist[12]).

Being pumped by a Coherent Verdi V-10 Neodymium:Vanadate (Nd:VO_4) solid state laser with a power of 10 W, the excitation source was derived from a Coherent Mira Ti:Sapphire femtosecond laser. The output wavelength of the latter ranges from 700 to 1000 nm. Nitrogen was used to keep low humidity in the laser cavity when operating at wavelengths above 900 nm.

The laser output was mode-locked and produced ~200 fs pulses at a 76 MHz repetition rate. A Coherent 9200 pulse-picker was used to reduce the repetition rate to 4.75 MHz, such that a pulse is emitted every 210 ns. The laser was then directed into a Coherent 5-050 harmonic generator where the laser frequency doubling or tripling can be done in order to obtain the desired excitation wavelength. A fraction the laser output from the harmonic generator was diverted to photodiode, which converted the laser pulse to an electronic pulse, which can provide *START* signal for the time-to-amplitude converter (TAC). The amplitude of the photodiode output pulse was set to 0.7~0.8 V. The intensity of the excitation beam can be reduced with neutral density filters. The excitation beam was vertically polarised using a Soleil-Babinet compensator, and then directed into the sample chamber and loosely focused on the sample holder of Edinburgh Instruments spectrometer. The position of Soleil Babinet Compensator, which modifies the orientation of linear polarised laser light must be tuned appropriately according to different excitation wavelengths. This required a resolution values to be set in the Edinburgh Instruments F900 control software. Equation 3.1 shows the relationship between resolution value (R) and wavelength λ.

Equation 2.4

$$R = \frac{41.62}{\lambda}$$

By using the card-based system EI TCC900 photon counting electronics together with an Edinburgh Instruments spectrometer and EI F900 software, fluorescence decays were measured. The *start* signal for the time-to-amplitude converter (TAC) was provided by the photodiode triggered by the residual laser from the harmonic generator, the fluorescence emitted from the sample after excitation provided the *stop* signal. Fluorescence emission of the sample was directed through a polariser set at the magic angle (54.8^{o}) to avoid the anisotropic effect on the results and detected thereafter by a Hammamatsu R3809 microchannel plate photomultiplier tube. Before detection the fluorescence was dispersed by a monochromator, which was set for a wavelength resolution of 10-15 nm, depending on the emission intensity. Any

scattered light from the excitation beam was blocked with a Schott longpass filter selected for the appropriate wavelength, located immediately after the cuvette.

The wavelength of the excitation was checked by conducting the steady-state emission scan on the unpolarised excitation beam. The alignment of the system was optimised by measuring the count rate of a 5 μM free 2AP solution as a standard solution. An instrument response function (IRF) was recorded using Ludox solution to scatter the excitation laser beam. The IRF for this system had a ~80 ps full width at half maximum intensity (FWHM) as shown in Figure 2.5. IRF shows the spectrometer's response to the excitation pulse. The FWHM of the IRF is mainly determined by the transit time spread for the MCP. However, FWHM is also affected by a few factors including the physical properties of the sample, the excitation and emission wavelengths used and the settings of the counting electronics of the spectrometer. It is crucial to record the IRF for decay fitting owing to the difference between the recorded width of the excitation pulse and the width of the pulse incident on the sample. This effect must be taken into account during the decay fitting process as described below.

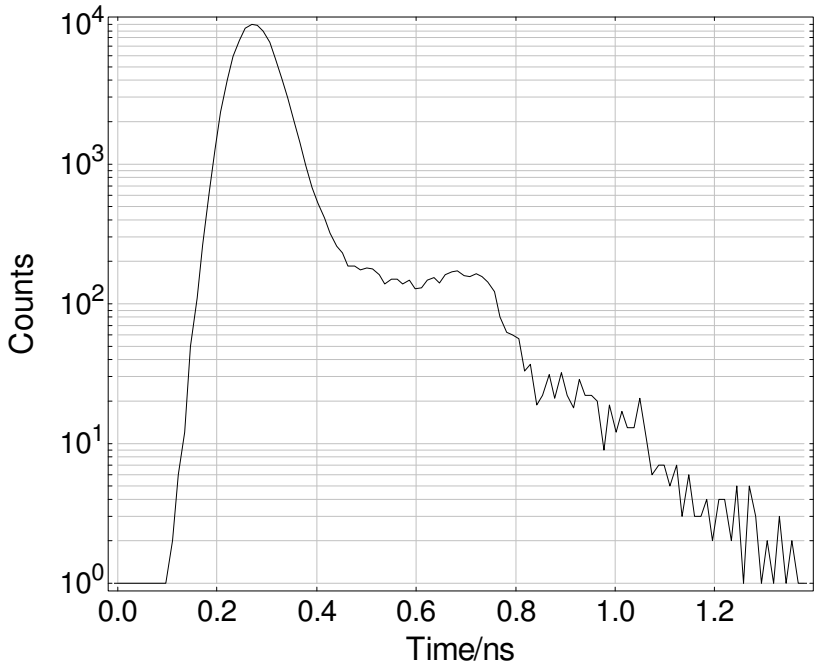


Figure 2.5—An example of instrument response function with ~85 ps FWHM. The Y-axis is shown on logarithmic scale.

Samples in aqueous solution were measured within a Starna fused silica micro-cuvette of 100 µL capacity, having a 2×1 mm excitation window and a 2×10 mm emission window. In general, a 1 µM concentration of the 2AP-labelled DNA was used for the DNA-enzyme experiments otherwise specified. 5-10 µM 2AP-labelled DNA was used for free oligonucleotides measurements.

An example of typical fluorescence lifetime decay curve obtained using TCSPC techniques is shown in Figure 2.6. Individual decays were analysed by the Edinburgh Instruments F900 software, which adopts *non-linear least squares* curve fitting approach.

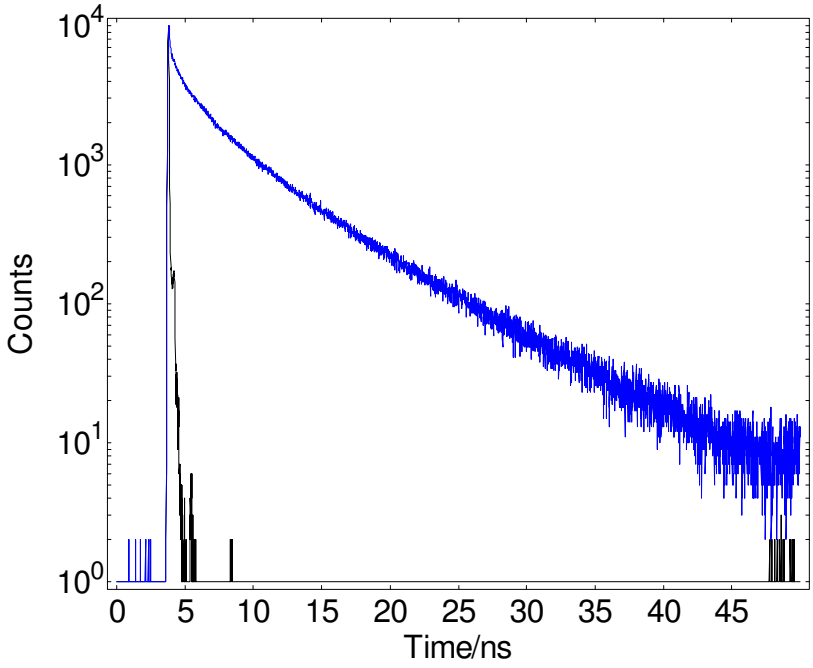


Figure 2.6—Fluorescence decay curves of 2AP-labelled DNA collected using the time-correlated single photon counting technique coloured in blue. The instrument response function (IRF) separately measured and coloured in black.

Fluorescence decay can be readily fitted using the equation as follows:

Equation 2.5

$$I(t) = B + \sum_{i=1}^{n} A_i \exp(\frac{-t}{\tau_i})$$

where B represents a constant accounting for any background intensity , A_i is the amplitude for the i^{th} decay component and τ_i is the fluorescence lifetime of the ith decay component. Equation 2.5 fails to take the instrument response function into account when fitting the decay data. The observed decay, F(t), is a convolution of the instrument response function, E(t), and the true decay of the 2AP, I(t):

Equation 2.6

$$F(t) = E(t) \otimes I(t)$$

Where $\otimes$ is the convolution operation, and I(t) is obtained by iterative reconvolution and uses a previously recorded instrument response function (IRF) as E(t). When fitting a decay curves, individual τ needs to be determined, which can be done through the software by minimising the weighted sum of squares of residuals between the fitted curves and the actual experimental curves. This operation is base on the equation as follows:

Equation 2.7

$$\chi_{\alpha}^{2} = \sum_{i=n_1}^{n_2} \left\{ \frac{\left[I_0(t_i) - Y(t_i) \right]^2}{I(t_i)} \right\}$$

where I0(t) is the background-corrected intensity decay, Y(t) is the fitted curve function and I(t) is the uncorrected intensity function. n1 and n2 are the first and last data point in the region of fitting respectively

The quality of any given fit is judged from the value of the reduced chi-squared function, χ^2

Equation 2.8

$$\chi^2 = \frac{\chi_a^2}{n_2 - n_1 + 1 - p}$$

where p is the number of variable parameters in the fitting function. χ^2 represents how good the fitting is; a value of 1 indicates a perfect fit has been accomplished. Values that deviate from this can be the results of either a poor fit or of 'imperfect'

data. For the all the decays presented in this thesis, a value for χ^2 of less than 1.2 typically signifies a 'good fit'.

Equation 2.9

$$\chi^2 = \frac{\chi_\alpha^2}{n_2 - n_1 + 1 - p}$$

The goodness of the fitted decay function can be further assessed by the inspection of a plot of the weighted residual $r(t_i)$ for each data point.

Equation 2.10

$$r(t_i) = \frac{I_0(t_i) - Y(t_i)}{\sqrt{I(t_i)}}$$

A good fit should have plotted residuals which are randomly scattered around zero. Decay curves were fitted based on Equation 2.5, which uses the smallest number of exponential decay terms that is able to give a reasonable χ^2 value and a good visual fit with randomly varying residuals. Addition of exponential terms is only to be considered if they significantly reduce the value of χ^2. The fitting range of a decay starts from the rising edge of the decay curve and ends in the point of the decay where the photon counts for that channel were approximately 10 times the background level. The background values used in fitting can be estimated from the number of counts arriving before the rising edge of the fluorescence decay and were fixed for the fitting procedure. The cooled MCP detector used in these experiments gives extremely low background counts, typically only 1-2 counts for a decay with a maximum intensity of 10,000 counts.

For example, Figure 2.7 shows the decay fitted with three exponentials using in Equation 2.5. Figure 2.8 shows the same decay fitted with four exponentials. The result of introducing a fourth exponential term gave rise to a significantly lower χ^2, a more smooth variation of the residuals regarding the actual experimental data. Thus the decay is best described by a four-exponential function as shown in Figure 2.8.

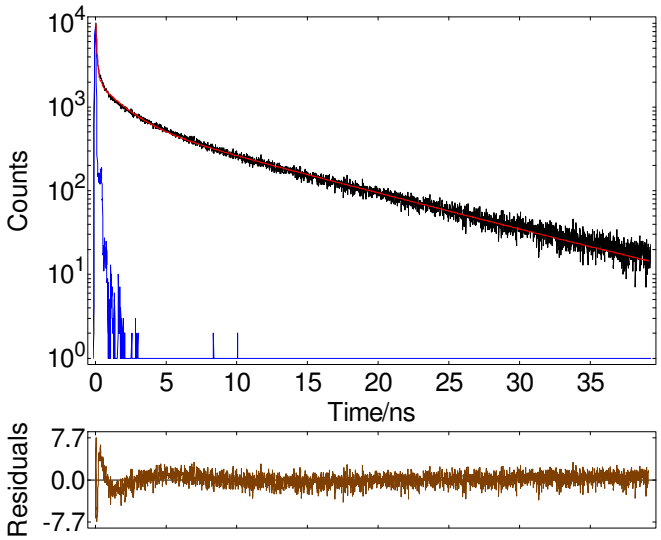


Figure 2.7—An example of typical fluorescence decay for 2AP-labelled dsDNA (black), the fitted decay generated by F900 (red) and the instrument response function (blue). A plot of the residuals between the decay data and the calculated function is also shown (brown). The decay was fit with three exponentials between the 297th and the 3800th time channel, and using a background level as 1 count. The resulting lifetimes calculated for the red function are 0.07, 1.77 and 9.988 ns and the reduced chi squared=1.506.

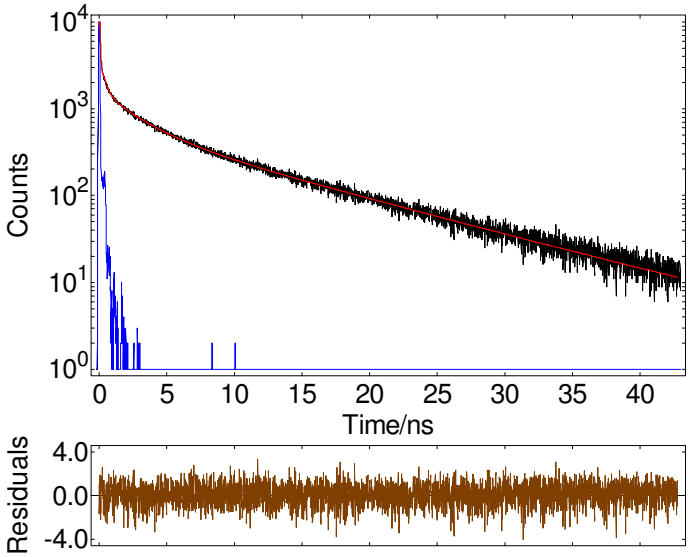


Figure 2.8—An example of fluorescence decay for 2AP-labelled dsDNA (black), the calculated function generated by F900 (red) and the instrument response function (blue). A plot of the residuals between the decay data and the calculated function is also shown (brown). The decay was fit with four exponentials between the 297th and the 3800th time channel, and using a background level as 1 count. The resulting lifetimes calculated for the red function are 10.4, 2.57, 0.33 and 0.04 ns, and the reduced chi squared=1.071.

For a given sample, decays were collected at three independent emission wavelengths (typically 370, 380 and 390 nm for the nucleic acid and 365, 380, 395 nm for enzyme-DNA experiments). Decays were fitted individually with F900 followed by a global fitting to generate the universal values of lifetimes (τ_i) and wavelength-specific fractional amplitudes and Chi-square for each emission wavelengths. Global fitting was carried out using Edinburgh Instruments FAST software.

The decay parameters derived from the fitting based on Equation 2.5 are τ_i, the fluorescence lifetime of the i^{th} emitting species, and A_i, the corresponding fractional amplitude of the i^{th} emitting species. Other useful quantities that can be calculated using these decay parameters, some of which are presented in the results chapters to follow, are the quantum yield of the sample relative to free 2AP riboside in solution, Φ_{rel}, (τ_{2APr}=10.6 ns):

Equation 2.11

$$\Phi_{rel} = \frac{\sum_{i=1}^{n} A_i \tau_i}{\tau_{2A\mathrm{Pr}}}$$

the contribution of the n^{th} emitting species to the steady state intensity (for a multi-exponential decay):

Equation 2.12

$$\frac{A_n \tau_n}{\sum_{i=1}^{n} A_i \tau_i}$$

the number-averaged lifetime of the sample:

Equation 2.13

$$\frac{\sum_{i=1}^{n} A \tau_i}{\sum_{i=1}^{n} A_i}$$

2.3 References

1. C. R. Cantor, M. M. Warshaw and H. Shapiro, *Biopolymers*, 1970, **9**, 1059-1077.
2. J. H. Murphy and T. L. Trapane, *Anal. Biochem.*, 1996, **240**, 273-282.
3. G. Kallansrud and B. Ward, *Anal. Biochem.*, 1996, **236**, 134-138.
4. I. Sissoeff, J. Grisvard and E. Guille, *Prog. Biophys. Mol. Biol.*, 1976, **31**, 165-199.
5. R. Owczarzy, A. V. Tataurov, Y. Wu, J. A. Manthey, K. A. McQuisten, H. G. Almabrazi, K. F. Pedersen, Y. Lin, J. Garretson, N. O. McEntaggart, C. A. Sailor, R. B. Dawson and A. S. Peek, *Nucleic Acids Res.*, 2008, **36**, W163-W169.
6. J. F. Sambrook, D. W. Russell and Editors, *Molecular cloning: A laboratory manual, third edition*, Cold Spring Harbor Laboratory Press, 2000.
7. S. C. Gill and P. H. von Hippel, *Anal. Biochem.*, 1989, **182**, 319-326.
8. C. N. Pace, F. Vajdos, L. Fee, G. Grimsley and T. Gray, *Protein Sci.*, 1995, **4**, 2411-2423.
9. T. J. Su, B. A. Connolly, C. Darlington, R. Mallin and D. T. F. Dryden, *Nucleic Acids Res.*, 2004, **32**, 2223-2230.
10. J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Plenum Press, 1983.
11. J. R. Lundblad, M. Laurance and R. H. Goodman, *Mol. Endocrinol.*, 1996, **10**, 607-612.
12. E. Y. M. Bonnist, *PhD thesis, The University of Edinburgh*, 2008.

Chapter 3---Local Structure of a DNA Three-Way Junction (3WJ) Revealed by Time-Resolved Fluorescence of 2AP

3.1 Introduction

3.1.1 Three-way junctions and their importance in biology and chemistry

Branched nucleic acids can be defined as molecules in which a number of distinct helical segments, with or without the intervention of formally single-stranded regions, are joined through the covalent continuity of the constituent strands.[1] Figure 3.1 shows the subtypes of DNA. Branched nucleic acids can be widely found in the biological world, and are deemed as important non-helical structures in many circumstances, playing a very crucial role as intermediates in genetic rearrangement processes such as replication, recombination, transposition, repair and integration events.[2-4]

As one of the important branched nucleic acid structures, 3WJs can be divided into two types, the perfectly base-paired 3WJ, and the bulged 3WJ. The former comprises three helices linked through the covalent continuity of the strand, with no unpaired bases; while the latter, has one or more extra base which is free from pairing with the complementary base(s) as shown in the Figure 3.1.[5,6]

For example, the 3WJ was found in the formation of very fast sedimenting DNA in genetic recombination of phage DNA and exists as structural elements in the terminal repeats of certain viruses.[7] It can be used as a simple model for studying the structure and dynamics of branched DNA in biology, such as the DNA replication fork. The 3WJ also plays a role in some DNA repeats capable of forming alternative structures from the canonical B-form helix. An illustrative example would be $(CAG)_n$•$(CTG)_n$ trinucleotide repeats, which are related to some human neurological diseases, such as HD.[8,9] It is believed that CAG and CTG repeats are able to form an unusual 3WJ structure as shown in Figure 3.2. It has been suggested that unusual slipped DNA structures caused by triplet repeats decrease the transcriptional efficiency *in vitro*. The possibility has been raised that if these structures occur *in vivo*, they may be involved in the aetiology of CAG repeat diseases like HD.[10]

3WJ structures occur frequently in functional RNA molecules as well. The hammerhead ribozyme is a notable example of a catalytic RNA that is built around a 3WJ.[11,12] Certain kinds of riboswitches also adopt a three-helical fold that centres on a 3WJ motif.[13]

Additionally, there is fast-expanding interest in the application of 3WJ in the research areas such as fluorescence detection, imaging technologies, building artificial multi-component nano-machines, and delivery of therapeutics.[14-19] It has been shown that a 3WJ could be used as an appropriate target for the design of new ligands with a high structural specificity and strong-binding affinity, and potentially be a drug target like the G-quadruplex.[20,21] Its geometry is also being exploited for building new supramolecular structures.[22]

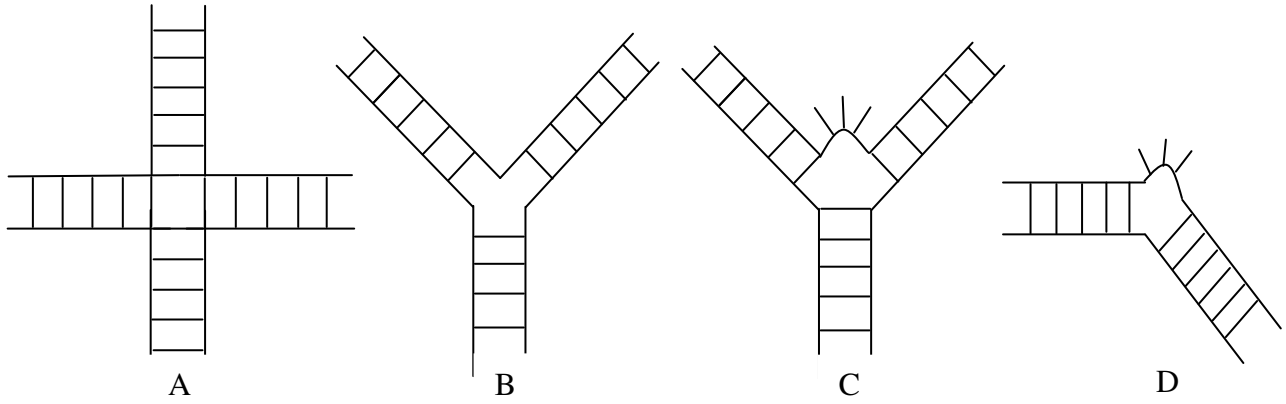


Figure 3.1—Some branched DNA species (adapted from Lilley *et al.*[1]). (A) Four way-junction which is stereochemically equivalent to the Holliday intermediate of recombination events. (B) Perfectly base-paired 3WJ. (C) Bulged 3WJ having unpaired bases in the branch point. (D) A bulge.

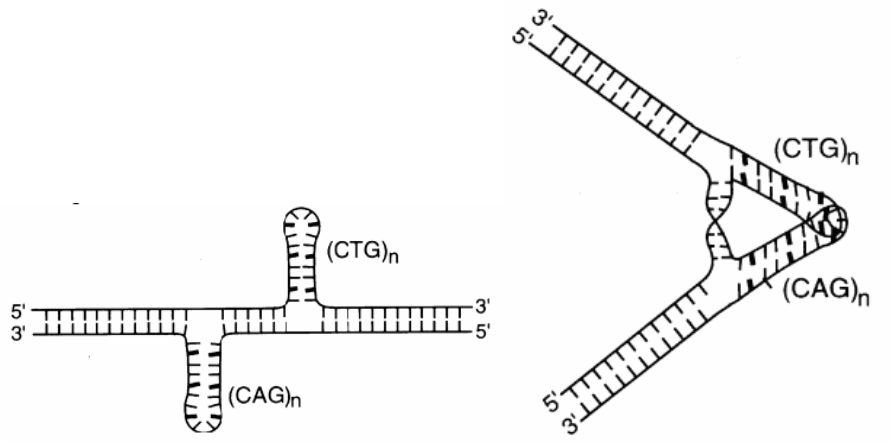


Figure 3.2—CAG and CTG repeats cause alternative DNA structures creating 3WJ (reprinted from Sinden *et al.*[9]). Left: slipped strand structure. Right: folded slipped strand stucture.

3.1.2 Previous studies of the global and local structure of 3WJs

It is can be envisaged that a perfectly based-paired 3WJ can adopt two structures: Y-structure or T-structure, as suggested by Duckett and Lilley[23] as shown in Figure 3.3. Welch *et al.*[24] used gel electrophoresis to investigate the geometry of 3WJ. The similarity in inter-arm angles between the arms, together with the reactivity of junction thymines against the osmium tetroxide modification, strongly suggested that the Y-shaped structure is the closer approximation to the 3WJ.

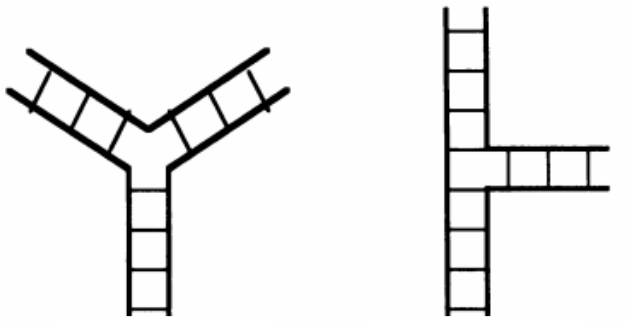


Figure 3.3—Two hypothetical structures of 3WJ in solution (reprinted from Duckett and lilley[23]). In the Y-shaped structure there is no helix-helix stacking, while in the T-shaped structure two of the helices undergo coaxial helical stacking forming a quasi-continuous helix.

Welch *et al.*[24] studied a series of DNA 3WJs containing different numbers of unpaired bases (0, 1, 2, 3 or 5) on one strand at the branch point (Figure 3.4A) and analysed their global conformations by polyacrylamide gel electrophoresis (PAGE). The sequences of a range of 3WJs were deliberately designed with some restriction enzyme cutting sites located in the arms, such that EcoRI, Xbal, and HindIII could cut specifically, resulting in three geometrically distinct fragments, as shown in Figure 3.4B. These cutting products were analysed using PAGE and the mobility of different 3WJ systems was revealed as shown in Figure 3.4D. These PAGE results were interpreted to propose that in the absence of an unpaired base, the 3WJ adopts an extended structure; in the junction having no unpaired base the three inter-arm angles are basically equivalent. However, with the addition of the unpaired bases in the branch point, the angle between the R and X arms increasingly widens as illustrated in the Figure 3.4E.

In a bid to see how metal ions affect the structural dynamics of 3WJ, the retardation of different 3WJs in the PAGE (Figure 3.5) was investigated in the presence of added magnesium ions. A model can be deduced from these gels: in the absence of unpaired bases, the mobility of the three species with one shorter arm is quite similar, inferring a structure in which the three inter-arm angles are quite similar. When inducing extra bases on the r strand the mobility of different species becomes very different, with HX being the fastest species. It is sensible to speculate that this is due to coaxial stacking of the H and X arms. The mobility of the RX species depends strongly on the number of added unpaired bases, which is in agreement with the angles between the R arm and the coaxial HX arms changing with the bulge size (number of unpaired bases) as shown in Figure 3.5. By comparing the data from Figure 3.4 and Figure 3.5, it is reasonable to construe that bulged 3WJs but perfectly base-paired 3WJs tend to undergo and cation ions-induced conformational change as illustrated in Figure 3.6.

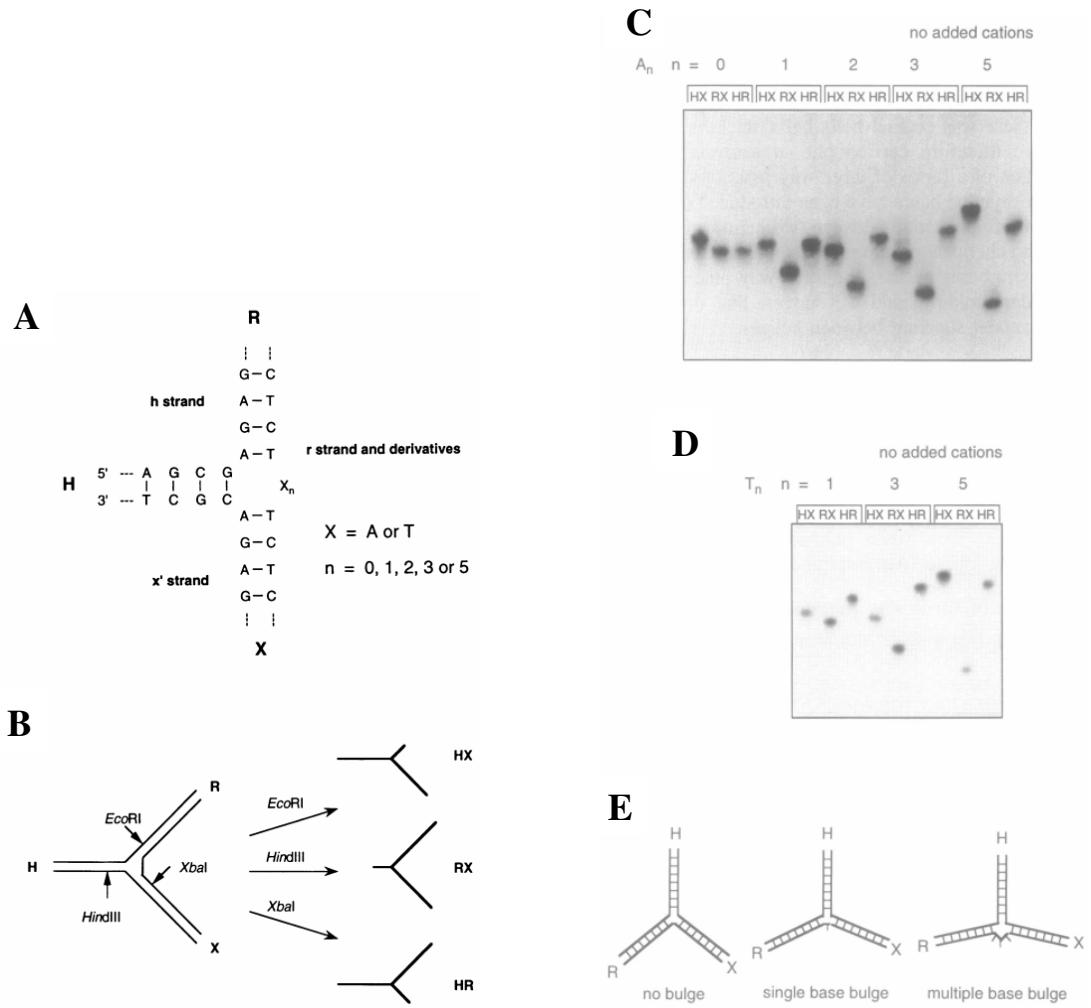


Figure 3.4—Gel electrophoresis method to study the structure of 3WJ (reprinted from Welch *et al.*[24]). (A) depicts the middle sequence of a series of 3WJs of interest. Each ssDNA composing 3WJs is labelled as h, r and x whose length is 80 mer (full sequences are not given), and thus the three arms are each 40 bp in length. Helical arms of 3WJs are named as H, R and X, each of which contains a unique site for a restriction enzyme (HindIII, EcoRI and XbaI) digestion respectively, allowing the arm to be selectively shortened by cleavage. (B) illustrates the structure of 3WJs analysed by PAGE. The junction is subjected to cleavage of one arm by HindIII, EcoRI and XbaI in separate digestion, and the electrophoretic mobility of the resulting species with two long and one short arm can thus be revealed for understanding the structure dynamics of 3WJ. These species are defined by the names of the two long arms (eg. HR has been cleaved with XbaI). (C) (D) (E) show the gel electrophoretic analysis of 3WJ in the absence of added metal ions. The three species with one shortened arm were examined by 8% polyacrylamide gel at room temperature in 90 mM Tris-borate, pH 8.3, 10 mM EDTA. (C) shows a series of 3WJs containing 0, 1, 2, 3 or 5 extra adenine bases on the r strand. (D) shows a series of 3WJs containing 1, 3 or 5 extra thymine bases on the r strand. (E) an illustration of the electrophoresis data shown. It can be proposed that

in the absence of added ions the junction adopts an extended structure. In the junctions having no unpaired bases the three angles between each arm are roughly equal; but on addition of bases on the r strand the angle subtended between the R and X arms increasingly widens.

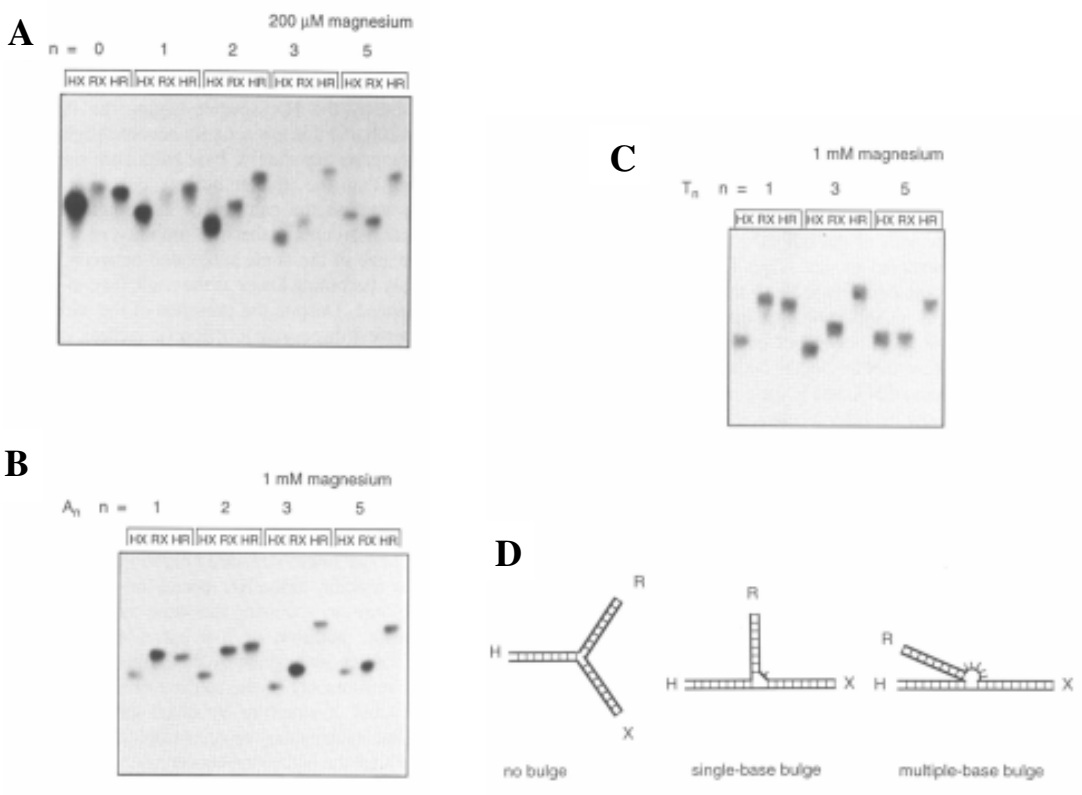


Figure 3.5—Gel electrophoresis study of 3WJs in the presence of added magnesium ions (reprinted from Welch *et al.*[24]). The three species with one shortened arm by electrophoresis in 8% polyacrylamide at room temperature in 90 mM Tris-borate, pH 8.3 with added $MgCl_2$. (A) a series of 3WJs containing 0, 1, 2, 3 or 5 extra adenine bases on the r strand, were electrophoresed in the presence of 200 µM magnesium ions. It is noticeable the difference in the relative mobility compared to the equivalent experiment in the absence of added magnesium ions (Figure 3.4B). (B) a series of 3WJs containing 1, 2, 3 or 5 extra adenine bases on the r strand, electrophoresed in the presence of 1 mM magnesium ions. (C) a series of 3WJs containing 1, 3 or 5 extra thymine bases on the r strand, electrophoresed in the presence of 1 mM magnesium ions. Comparison can be made regarding the relative mobility to its equivalent experiment in the absence of added magnesium ions (Figure 4C). (D) shows a structural model of 3WJ deduced from the gel data.

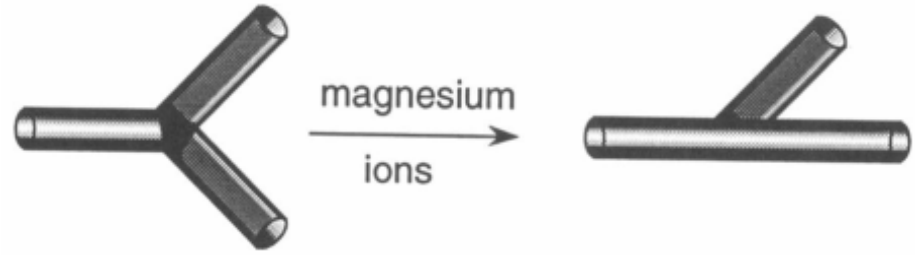


Figure 3.6—A proposed ion-dependent conformational transition model of bulged 3WJ induced by added magnesium (reprinted from Welch *et al.*[24]).

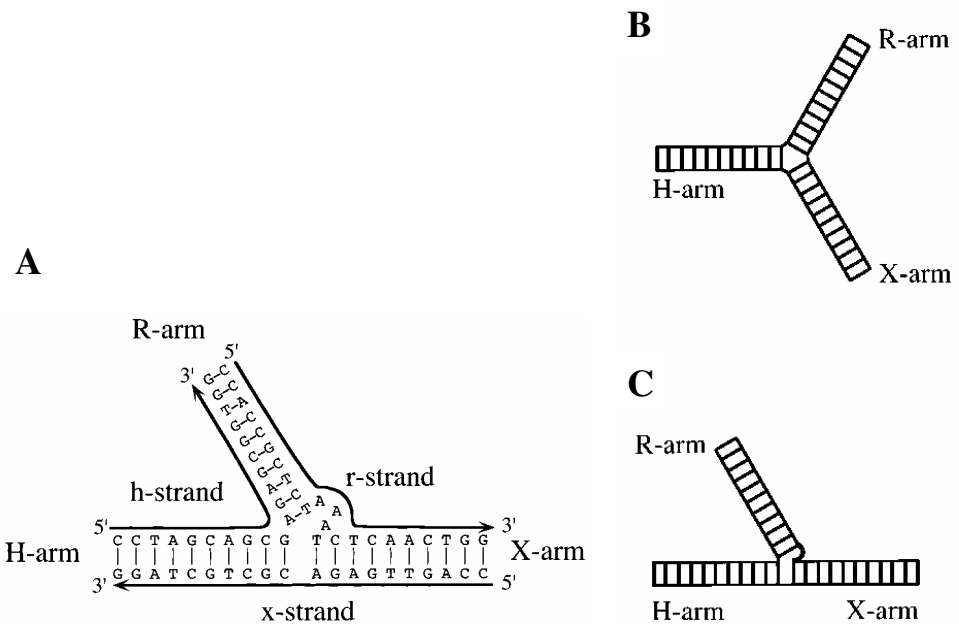


Figure 3.7—Schematic experimental design and explanation of the results derived from ensemble FRET measurements of 3WJs (reprinted from Stuehmeier *et al.*[27]). (A) 3WJ with three unpaired adenine nucleotides at the point of the strand exchange. The three arms are named as H, R, and X and the single strands are labelled h, r, and x. The arrows indicate 5'-3' direction of the strands. In each doubly labelled vector, the fluorescent donor and acceptor are covalently attached to the 5' ends of two of the three strands in each 3WJ molecule. (B) and (C) depict the global structures for the 3WJs with and without molecules bulge deduced from FRET data. (B) The model for the DNA 3WJ with no unpaired base at the branch point is displayed as a relatively symmetric Y-shaped molecule with approximately equal angles between the arms; there is no helical-helical stacking interaction between the arms even at higher salt concentration. (C) The unpaired adenine nucleotides in the r-strand at the branch point of the same molecule produce a coaxial helical stacking between two of the arms at higher salt concentrations hence being asymmetric. This structure represents a proposed stacking between the H and X arm; the R arm tilts away from the bulge.

Ensemble Förster resonance energy transfer (FRET), as a very useful technique for probing the structure dynamics has also been used for study the structure of 3WJ.[25,26] In the work done by Stuehmeier *et al.*,[27] they studied the 3WJ with and without unpaired adenine base in one strand at the branch point (eg. A_n bulges with n=0, 1, 2, and 3). The structure of 3WJ having no unpaired base was found to have a relatively much more symmetric trigonal geometry, while 3WJ with unpaired bases (bulged), the arrangement turns out to be asymmetrical. Their FRET results suggested a model of bulged junctions where the angle between two of the arms is notably smaller than it between the other two arms as shown in Figure 3.7. The acute angle becomes smaller with the increase of the nucleotides number in the bulge. It is worth noting that the three helices may not be constrained to adopt a common planar conformation. It is notable there may be a certain amount of flexibility in the structure, and the stacked helices may not be perfectly linear as well.

Yang and Millar[26] used time-resolved fluorescence resonance energy transfer to probe the dynamics of 3WJ, whose conclusion is shown in Figure 3.8. By using a Gaussian distribution model, the distributions of fluorescent donor-acceptor (D-A) distances presented between each pair of labelled helices were recovered from analysis of the donor's decay. They concluded that for the junction lacking unpaired bases, the three mean inter-arm distances calculated are quite equal, indicating an extended structure. Additionally, a relatively broad range of distances is present between each pair of helices, showing that the structure is flexible. The introduced unpaired bases cause the junction to fold into a different structure, with one inter-arm distance being shorter than the other two. The change in overall geometry of the junction seems to be principally as a result of the repositioning of one of the helices flanking the bulge. In the bulged 3WJ having unpaired thymine, cytosine, or adenine bases, the helical arm III appears to undergo the greatest change with respect to the mean position relative to the other helices. On the contrary, in the bulged junction having unpaired guanine bases, the helical arm II is inclined relative to the other helices. Overall speaking, in all bulged 3WJs, there is a wide range of distances between the perturbed helix and the other two helices, indicative of high mobility for the perturbed helical arm. These results indicate that the overall structure and

conformational flexibility of DNA 3WJs are subjected to the presence of unpaired bases at the branch point of the junction and that the precise conformation of a bulged 3WJ highly depends on the nature of the unpaired bases.

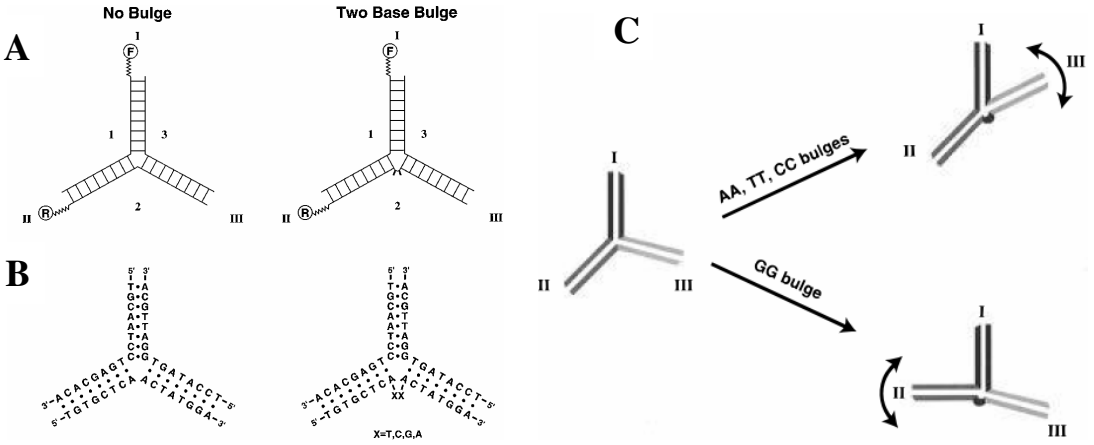


Figure 3.8—Schematic experimental design and explanation of the results derived from ensemble FRET measurement of 3WJ with and without added unpaired bases (reprinted from Yang and Millar[26]). (A) represents dye-labelled 3WJs without unpaired bases (left) and with two added unpaired bases in one strand (right-hand side). (B) Sequences of 3WJs without added bases (left) or with different type of added bases in strand 2 (right). (C) shows illustration of data obtained from FRET experiments. The global structure of the junction without added bases is relatively asymmetric, with the distance from arm I to arm II is higher than the other two inter-arm distances. When introducing two extra thymine, cytosine, or adenine bases at the branch point of the junction exhibits arm III to move closer to helix I and further away from helix II, while this has little effect in terms of the distance between helices I and II. Moreover, helix III has greater mobility in the bulged junctions and can be flexible in a wide range of positions, as indicated by the curved arrow. In contrast, when adding two extra guanines to the 3WJ, it causes arm II to move closer to arm I, while having little effect on the distance between arm I and III. Additionally, arm II has greater mobility in the bulged junction, again as indicated by the curved arrow.

Some endeavours have been made to determine the structure and dynamics of 3WJs using NMR spectroscopy. There are some reports of NMR-based structures of bulged 3WJs in solution but no report for fully base paired 3WJs so far. Thiviyanathan *et al.*[28] found that a DNA 3WJ containing two unpaired bases in the branch point of the junction exists in an extended conformation with two of the helical arms stacked

upon each other as shown in Figure 3.9A. The angles of each helical arm are quite similar and the two unpaired bases are exposed to the solvent. Helical parameters for the bases in all three strands show slight variations from typical values expected for canonical right handed B-form DNA. Wu *et al.*[29] proved that a 3WJ having two unpaired bases in one arm is able to adopt a coaxial helical-stacked conformation as shown in Figure 3.9B. This finding is highly in agreement with the previous studies using gel electrophoresis and ensemble FRET methods.

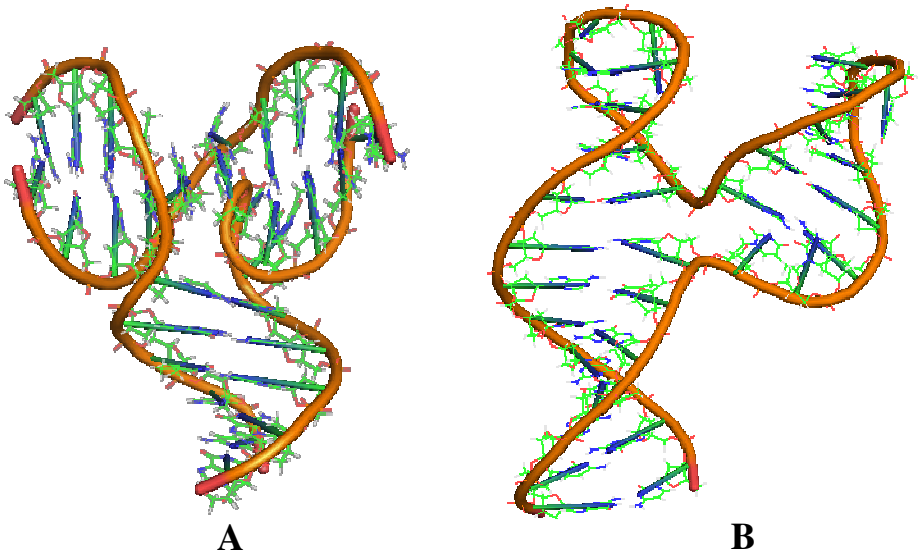


Figure 3.9—NMR-based structure of 3WJs. (A) 3D NOESY-NOESY NMR-based structure of the DNA 3WJ having two unpaired bases in one strand in the branch point of the junction (PDB: 1EKW). (B) NMR-based solution structure of the DNA 3WJ presenting a coaxial helical-stacked conformation. (PDB: 1SNJ)

In addition, AFM has been employed to investigate the global structure and conformation of 3WJs.[30] In the work of Shlyakhtenko *et al.*,[31] 3WJs containing one hairpin arm (50, 27, 18, 7 bp long) were examined using AFM as shown in Figure 3.10. AFM data (Figure 3.11 and Table 3.1) shows no dependence of the kink angle on the hairpin length. Besides, two characteristics of 3WJ were identified. Firstly, 3WJ has a very dynamic structure, which can be seen by a high deviation of its inter-arm angles measured on dried samples as well as the mobility of it directly imaged *in situ*. Secondly, the deduction from angles measured between different arms led to the conclusion that the geometry adopted by 3WJ is pyramidal rather than flat.

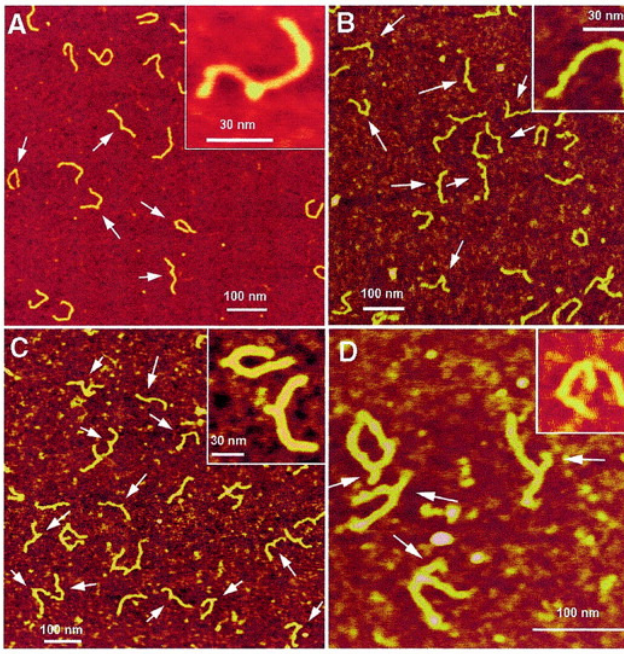


Figure 3.10—Image obtained of DNA 3WJs with different lengths of hairpins by AFM (reprinted from Shlyakhtenko *et al.*[31]). (A) 7 bp; (B) 18 bp; (C) 27 bp; (D) 50 bp. The hairpins are indicated by arrows. The sample was deposited on AP mica, argon dried and imaged in air under ambient conditions.

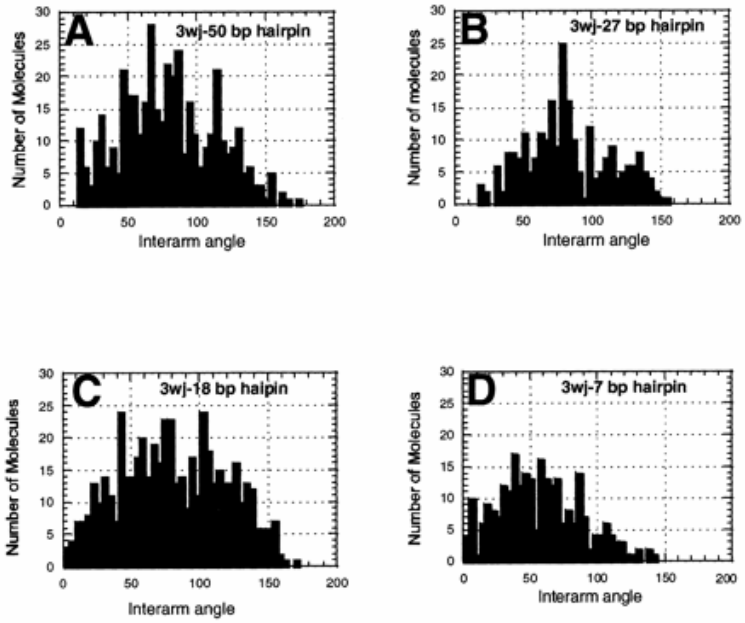


Figure 3. 11—The inter-arm angles of 3WJs containing hairpin with different lengths (reprinted from Shlyakhtenko *et al.*[31]). (A) 50 bp hairpin junction. (B) 27 bp hairpin junction. (C) 18 bp hairpin junction (D) 7 bp hairpin junction. The value in average and standard deviation are all listed in Table 3.1.

	Type of junctions (Number of molecules measured)			
	50 bp hairpin (416)	27 bp hairpin (240)	18 bp hairpin (507)	7 bp hairpin (256)
Mean angle between arms (°)	78±35	82±32	83±33	57±31

Table 3.1—The mean angles between linear arms for all 3WJs studied (adapted from Shlyakhtenko *et al.*[31]).

3.1.3 Single-molecule FRET together with molecular dynamics study of 3WJ

As a complementary tool to conventional methods, such as crystallography and NMR, single-molecule Förster resonance energy transfer (SM-FRET) is a fast expanding quantitative tool for elucidation of structural dynamics.[32,33] SM-FRET probes inter- or intra-molecular energy transfer between donor and acceptor dyes on the scale of 1-10 nm length, making it an ideal 'molecular ruler' for studying the structure and dynamics of nucleic acids.[34-36] However, the measurement of absolute distances requires knowledge of the relative orientations and mobility of the donor and acceptor dyes. Sabir *et al*.[37] used multi-parameter fluorescence detection (MFD), which allowed them to record the fluorescence intensity, colour, lifetime and polarization simultaneously. Also FRET-related structural dynamics was considered. The absolute donor-acceptor distances can be therefore worked out via a non-linear conversion function that takes into account the positional and orientational averaging of the dyes. Subsequently, the dye-dye distances were used as restraints in molecular dynamics (MD) simulations, as the helical arms of the structures were modelled and restrained as B-DNA. A combination of multi-parameter SM-FRET and molecular dynamics (MD) produced a 3D global structure of a forked DNA molecule in solution[37]. In this four-stranded fork structure, no evidence for unzipping at the branch point was identified, even when two base pairs at the branch were unrestrained. Recently, Sabir *et al*.[38] applied this technique to study the structure of 3WJ. The global structure of perfectly base-paired 3WJ was studied using

single-molecule FRET, generating a 3D picture of it in solution, as show in Figure 3.12 and Figure 3.13. Figure 3.12 illustrates the methodology of single-molecule MFD by which the absolute distance between each donor-acceptor pair can be measured. By placing different donor-acceptor pairs in different location of 3WJ, a series of distances can be obtained. Figure 3.13 shows the global structure of a 3WJ derived from SM-FRET distance restraints using the MD technique. It was found that the perfectly base-paired 3WJ adopts a Y-shaped, pyramidal structure and the bases adjacent to the branch point are unzipped, in spite of the full Watson-Crick complementarities of the nucleotides.[38]

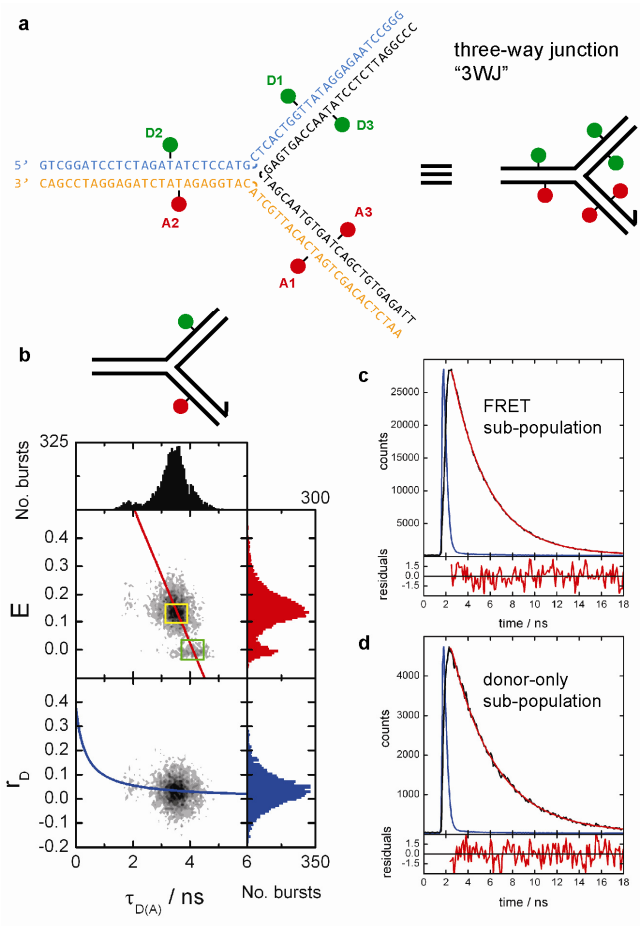


Figure 3.12—Single-molecule FRET for global structure elucidation of a 3WJ (reprinted from Sabir *et al.*[38]). (a) DNA sequence and position of donor (D) and acceptor (A) dyes. A simplified representation is depicted on the right. (b) Typical MFD data are shown for a three-way junction labelled donor and acceptor dyes. The 2D plots are of FRET efficiency (E) or donor anisotropy (r_D) versus donor lifetime ($\tau_{D(A)}$). The grey scale indicates an increasing number of single-molecule bursts (from white to black).

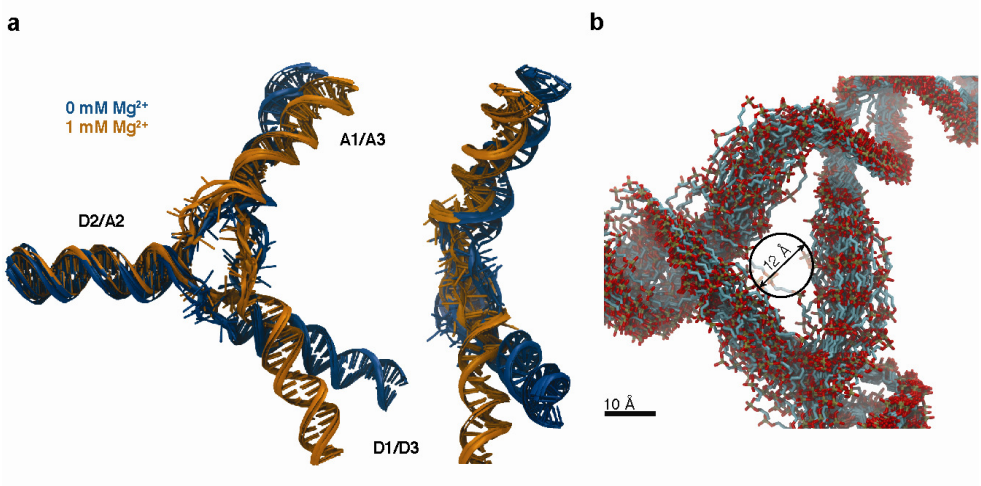


Figure 3.13—Global structure of 3WJ derived from SM-FRET distance restraints and MD simulations (reprinted from Sabir *et al.*[38]). (a) The five lowest energy structures are shown for 3WJ in buffer containing 0 mM $MgCl_2$ coloured in blue and 1 mM $MgCl_2$ coloured in orange. (b) Close-up view of the branch point region for a superposition of the 50 solution structures generated for 3WJ in 0 mM $MgCl_2$. The arrow indicates the diameter of the cavity accessible *via* unzipping of 2 base pairs in each arm.

3.1.4 Aims of this chapter

In spite of some previous work concerning the global structure of 3WJs, there appears to have been no comprehensive studies in which local conformational properties are also considered. In this chapter, the author would like to apply time-resolved fluorescence spectroscopy to study the local structural dynamics of 3WJs using 2AP-labelled oligonucleotides. Firstly, the ability of time-resolved fluorescence to reveal the unzipping of bases in DNA was examined. Secondly, a 3WJ was investigated to determine the nature and extent of conformational distortion compared with the equivalent duplex DNA. Time-resolved fluorescence measurements were made on a series of 3WJs, with 2AP at different locations relative to the junction but otherwise with identical sequences, along with the corresponding ssDNA and dsDNA. The sequences of the oligonucleotides and the locations of 2AP in the 3WJ are shown in Table 3.2. Comparison of the decay parameters for ssDNA and dsDNA enabled empirical indicators of unzipping to be established. These empirical indicators could then be used to investigate base unzipping in the 3WJ.

ssDNA1	GTCGGATCCTCT**2**GATATCTCCATGCTCACTGGTTATAGGAGAATCCGGG
ssDNA2	GTCGGATCCTCTAGATATCTCC**2**TGCTCACTGGTTATAGGAGAATCCGGG
ssDNA3	AATCTCACAGCTGATCACATTGCT**2**CATGGAGATATCTAGAGGATCCGAC
ssDNA4	AATCTCACAGCTGATC**2**CATTGCTACATGGAGATATCTAGAGGATCCGAC
ssDNA5	CCCGGATTCTCCT**2**TAACCAGTGAGTAGCAATGTGATCAGCTGTGAGATT
ssDNA6	CCCGGATTCTCCTATAACCAGTG**2**GTAGCAATGTGATCAGCTGTGAGATT
ssDNA7	CCCGGATTCTCCTATAACCAGTGAGTAGCAATGTGATC**2**GCTGTGAGATT

dsDNA1	GTCGGATCCTCT**2**GATATCTCCATGCTCACTGGTTATAGGAGAATCCGGG CAGCCTAGGAGATCTATAGAGGTACGAGTGACCAATATCCTCTTAGGCCC
dsDNA2	GTCGGATCCTCTAGATATCTCC**2**TGCTCACTGGTTATAGGAGAATCCGGG CAGCCTAGGAGATCTATAGAGGTACGAGTGACCAATATCCTCTTAGGCCC
dsDNA3	AATCTCACAGCTGATCACATTGCT**2**CATGGAGATATCTAGAGGATCCGAC TTAGAGTGTCGACTAGTGTAACGATGTACCTCTATAGATCTCCTAGGCTG
dsDNA4	AATCTCACAGCTGATC**2**CATTGCTACATGGAGATATCTAGAGGATCCGAC TTAGAGTGTCGACTAGTGTAACGATGTACCTCTATAGATCTCCTAGGCTG
dsDNA5	CCCGGATTCTCCT**2**TAACCAGTGAGTAGCAATGTGATCAGCTGTGAGATT GGGCCTAAGAGGATATTGGTCACTCATCGTTACACTAGTCGACACTCTAA
dsDNA6	CCCGGATTCTCCTATAACCAGTG**2**GTAGCAATGTGATCAGCTGTGAGATT GGGCCTAAGAGGATATTGGTCACTCATCGTTACACTAGTCGACACTCTAA
dsDNA7	CCCGGATTCTCCTATAACCAGTGAGTAGCAATGTGATC**2**GCTGTGAGATT GGGCCTAAGAGGATATTGGTCACTCATCGTTACACTAGTCGACACTCTAA

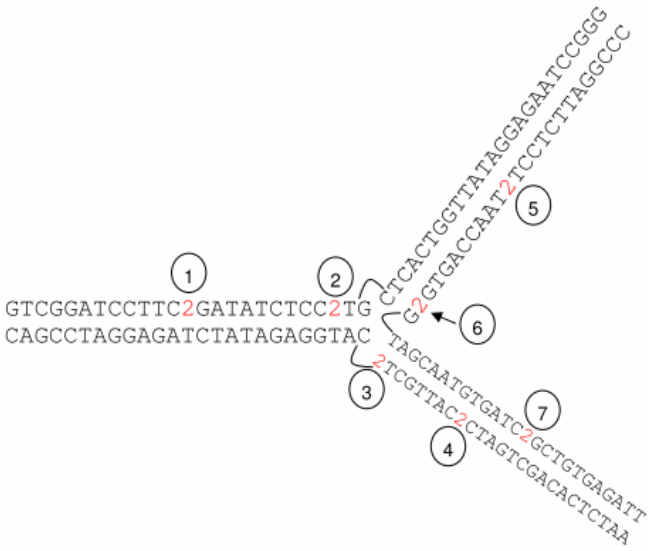


Table 3.2—Sequences of ssDNA, dsDNA and 3WJ samples used in the study. Note that 3WJ figure shows a composite of all 2AP positions. Four of the labels are within the long duplex region of an arm, while three of the labels are near the branch point. The numbers in the circles indicate the names of 3WJs with different 2AP-labelled positions. (2=2AP base).

3.2 Experimental

3.2.1 DNA sample preparation

2AP-labelled oligonucleotides were synthesised by Purimex GmbH (Grebenstein, Germany). Annealing of samples for branched DNA and dsDNA was carried out in buffer (20 mM Tris, 15 mM NaCl, pH 7.5). For all structures the ratio of 2AP-labele dstrand to unlabelled strand was 1:3. Samples were heated to 90°C in a water bath and left to cool slowly overnight.

3.2.2 Time-resolved fluorescence measurements

Time-resolved fluorescence was recorded using TCSPC as described in Chapter 2. The fluorescence lifetimes were measured at excitation wavelength 310 nm. The 310 nm excitation radiation was produced by frequency tripling the 930 nm output of the mode locked Ti: Sapphire laser. The fluorescence lifetime decays were collected at three emission wavelengths: 370 nm, 380 nm and 390 nm, respectively. The slits of monochromator were set to 10 nm bandpass. The concentration of ssDNA, dsDNA and 3WJ was 10, 5, 7 μM, respectively, and all were buffered with 20 mM Tris, 15 mM NaCl (pH 7.5). All the measurements were made at room temperature.

3.3 Results and Discussion

3.3.1 Decay parameters of the dsDNA samples

The fluorescence decay parameters for the dsDNA samples are shown in Table 3.3.

All of the seven duplexes show four discrete lifetimes, as reported previously for similar systems.[39-42] The shortest lifetime accounts for most of the population, but the A-factor varies significantly with sequence context, from 0.55 to 0.89. Taken together, the two shortest lifetimes account for the vast majority of the population, >0.85, in all cases. The longest lifetime represents a very small fraction of the

population (0.01 to 0.05). It is a constant trend that the A-factor decreases with increasing fluorescence lifetime for all the seven duplexes, which is in agreement with previous observations for 2AP-labelled dsDNA.[40,42]

dsDNA	A-factor				Lifetimes/ns			
	A_1	A_2	A_3	A_4	τ_1	τ_2	τ_3	τ_4
dsDNA1	0.70	0.18	0.09	0.04	0.08	0.50	2.89	10.2
dsDNA2	0.84	0.10	0.04	0.02	0.05	0.37	2.66	10.6
dsDNA3	0.89	0.08	0.02	0.01	0.10	0.35	2.36	9.70
dsDNA4	0.76	0.18	0.04	0.02	0.04	0.38	2.76	10.5
dsDNA5	0.53	0.41	0.04	0.01	0.08	0.35	2.51	9.10
dsDNA6	0.65	0.20	0.11	0.04	0.07	0.45	2.64	8.81
dsDNA7	0.70	0.17	0.08	0.05	0.06	0.44	2.75	10.8

Table 3.3—Results derived from global analysis of fluorescence decay curves for 2AP-labelled dsDNA samples in buffered aqueous solution. The A factors show little wavelength-dependence and the reported values are the average of measurements at 370, 380, 390 nm).

The effect of the sequence context on decay parameters is most evident with respect to the two shorter lifetimes and their respective amplitudes. The very short lifetime, τ_1, is the result of fast interbase charge transfer quenching. The rate of charge transfer to excited 2AP differs between the natural nucleobases; it is greatest for G and decreases with the following trend: G>T>A>C.[43,44] In all seven duplexes, there is at least one G in close proximity to 2AP, either in the same strand or in the opposite strand. This is consistent with the observation of a very short decay component of 100 ps or less in all cases. Looking closer, the shortest τ_1 (0.04 ns) belongs to dsDNA4, in which there are two guanines immediately adjacent in the opposite strand. Although dsDNA6 has a guanine on each side of 2AP in the same strand, 2AP shows a longer τ_1 at 0.07 ns, indicating that factors other than the proximity of G affect the rate of quenching. The longest τ_1 (0.10 ns) is found for dsDNA3 which 2AP has one adjacent G in the opposite strand on the 5' side. In comparison, dsDNA2 with an adjacent G in the opposite strand in the 3' side has a shorter τ_1 of 0.05 ns.

When considering the fractional amplitudes, what can be observed is that dsDNA3 (A_1=0.89) and dsDNA2 (A_1=0.84) display the highest two A_1 values, indicating the highest populations of a highly stacked conformation. This can be correlated with the proximity of multiple G:C base pairs (dsDNA2 has four and dsDNA3 has three G:C base pairs located very close to 2AP:T base pair). The A_1 values for the others, dsDNA1, 4, 6, 7 range from 0.65 to 0.76, which are very comparable. The A_1 and A_2 values for dsDNA5 are unusual, 0.53 and 0.41 respectively. However, this is in agreement with the previous study by Rai and Millar[43], who found that 2AP inside the ATAT (and TATA) duplex sequence shows unusual conformational behaviour. Bonnist *et al.*[39] also observed a very similar pattern for 2AP inside in an ATAT sequence.

For lifetime components 3 and 4, all the duplexes behave as expected, τ_3 for all duplexes fall in the time range from 2.36 to 2.89 ns; τ_4, descriptive of an extrahelical base exposed to a water-like environment, are all around 10 ns. The sum of A_3 and A_4 for each dsDNA is less than 0.15, A_4 values all fall range of 0.04 to 0.07; while A_3 values are around 0.07, which is very characteristic of 2AP-labeled duplexes in good agreement with numerous studies done by our group and others before.

3.3.2 Decay parameters of the ssDNA samples

For the ssDNA samples, the results show they all exist in multiple conformational states, manifested by a complex decay that can be described by the sum of four exponential components with typical lifetimes of <100 ps, 0.5 ns, ~2 ns, and ~10 ns as shown in Table 3.4.

ssDNA	A-Factor				Lifetimes/ns			
	A_1	A_2	A_3	A_4	τ_1	τ_2	τ_3	τ_4
ssDNA1	0.69	0.13	0.13	0.05	0.07	0.66	2.51	7.39
ssDNA2	0.67	0.16	0.14	0.04	0.07	0.58	2.21	6.75
ssDNA3	0.65	0.16	0.21	0.04	0.09	0.55	2.22	6.09
ssDNA4	0.57	0.20	0.17	0.06	0.10	0.63	2.54	7.37
ssDNA5	0.64	0.13	0.18	0.05	0.06	0.53	2.07	5.40
ssDNA6	0.64	0.14	0.15	0.07	0.07	0.69	2.87	7.60
ssDNA7	0.69	0.14	0.13	0.04	0.07	0.68	2.40	7.80

Table 3.4—Results derived from global analysis of fluorescence decay curves for 2AP-labelled ssDNA samples in buffered aqueous solution (Reported A-factors are the average values for 2AP emission at 370, 380, 390 nm).

It is obvious that the shortest lifetime, τ_1, persists for all ssDNA samples, signifying that the most highly stacked conformation persists in the absence of base pairing. However, unlike dsDNA, extremely short lifetime values, <70 ps, are not observed apart from ssDNA5. In all cases, τ_2 for ssDNA is significantly longer than in the corresponding duplex. The most highly stacked conformation still remains highly populated for ssDNA (60-70%), but the value of A_1 shows much less variation, indicating it has less dependence on sequence context than in duplexes. In contrast to τ_1, τ_4 is significantly shorter than in the duplex in all cases, which means 2AP does not completely escape from interbase interactions in the single strand, whereas in the duplex this component is typically characteristic of extrahelical (solvent-exposed) 2AP.

The most notable difference between duplex and single strand is the relative values of the A-factors. In the duplex (with the exception of TATA box, DNA5), A-factors follow the trend $A_1>>A_2>A_3>A_4$. In single strand, population is transferred from the highly stacked states (A_1 and A_2) to poorly stacked states (A_3), so that the trend becomes $A_1>>A_2 \approx A_3>>A_4$ as shown in Figure 3.14.

The extent to which decay parameters change between dsDNA and ssDNA depends

on the sequence context. For both DNA2 and DNA4 there is a marked increase in τ_1 for single strand compared with the duplex, clearly demonstrating the quenching effect of adjacent Gs in the complementary strand in these duplexes. Interestingly, this effect on τ_1 is not seen for DNA3 (ssDNA3, dsDNA3), suggesting that the 5' G in the complementary strand of dsDNA3 does not contribute significantly to quenching. For DNA1, DNA6 and DNA7, A_1 and τ_1 are almost unchanged between dsDNA and ssDNA, since each has a 3'G stacked with 2AP in the same strand. Duplexes with very high populations of highly stacked states (A_1>0.8), for instances dsDNA2 and dsDNA3, show a sizeable increase in A_1 than in the corresponding ssDNA samples, confirming the contribution of base pairing for maintaining these highly quenched states. These are the duplexes in which 2AP occurs in a G:C-rich region. The Anomalous A_1 and A_2 values for dsDNA5 (TATA box) are lost in ssDNA, which indicates that the unusual conformational properties of the TATA box arise from base pairing, not stacking.

3.3.3 Empirical indicators of unzipping

The intention here was to determine which lifetime parameters can help one to discern ssDNA and dsDNA or what are the most revealing differences between ssDNA and dsDNA in terms of lifetime components, in order to define some empirical indicators of base pairing/unzipping.

As shown in Figure 3.14, the plot of A-factor *vs.* τ gives a convenient way of representing the decay parameters and making the comparison between ssDNA and corresponding dsDNA. The difference in the shape of this plot, between dsDNA and ssDNA, can be seen in each case and the sequence-dependence difference is apparent.

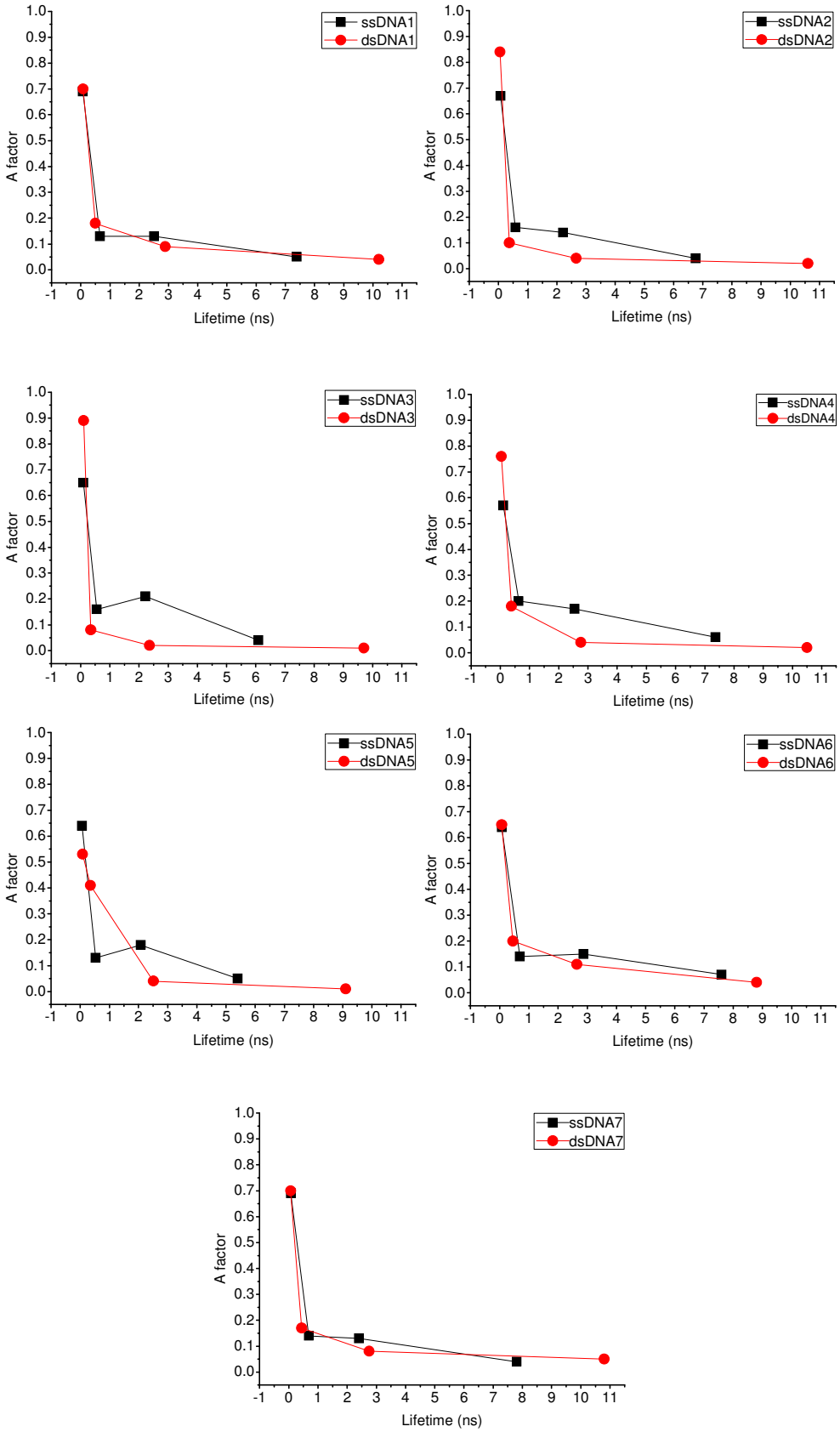


Figure 3.14—Plots of A-factor *vs.* lifetime for ssDNA and corresponding dsDNA samples.

The most characteristic observation is the change in the values of A_2 and A_3 between ssDNA and dsDNA, as illustrated in Figure 3.15. For dsDNA, A_2 is consistently much higher than A_3; however, for ssDNA it is invariably the tendency that A_2 decreases and A_3 increases significantly, resulting in A_2 and A_3 having more or less the same values. In some cases (ssDNA3, ssDNA5 and ssDNA6) A_3 is even higher than A_2. This signifies that the loss of hydrogen bonding in ssDNA, and the consequently greater freedom in base dynamics, is mainly reflected in the base conformations that give rise to τ_2 and τ_3. In general, the equilibrium of the conformational states shifts from a quite highly stacked conformation (represented by τ_2) to a less stacked conformation (represented by τ_3).

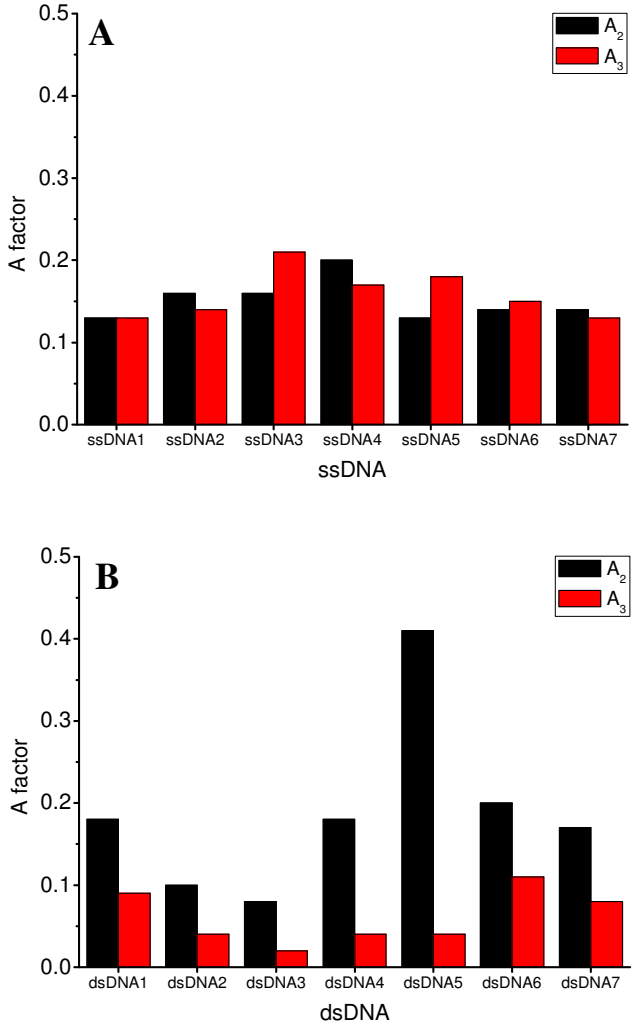
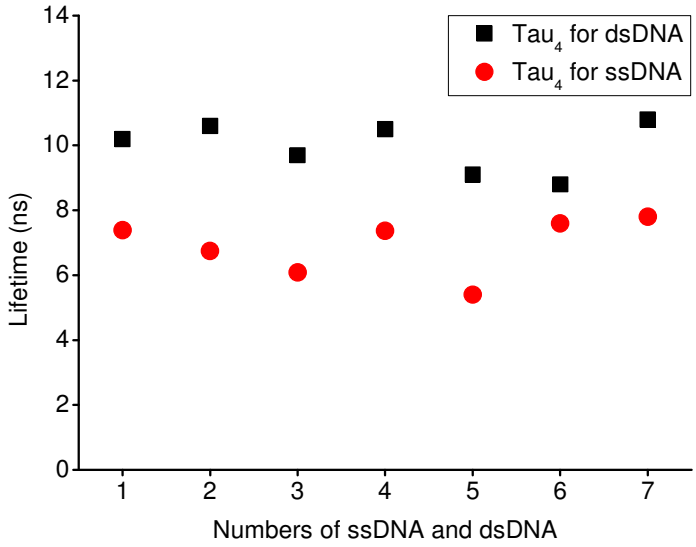


Figure 3.15—Comparison of A_2 and A_3 values for ssDNA and dsDNA samples. (A) The A_2 and A_3 values for ssDNA samples. (B) The A_2 and A_3 values for corresponding dsDNA samples.



	τ_4/ns						
Sample Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**
ssDNA	7.39	6.75	6.09	7.37	5.40	7.60	7.80
dsDNA	10.2	10.6	9.7	10.5	9.1	8.8	10.8

Figure 3.16—Comparison of τ_4 values for ssDNA and dsDNA samples.

Another consistent difference is in the value of τ_4, as shown in Figure 3.16. As noted before, the shortened τ_4 value for ssDNA indicates that 2AP does not completely escape from interbase interactions, while it does in the duplex where this component is very characteristic of extra-helical 2AP base.

3.3.4 Decay parameters and conformational properties of the 3WJ

Lifetime parameters for 2AP in the seven different positions in the DNA 3WJ are given in Table 3.5. All show some deviation from the corresponding dsDNA parameters, to some extent. It can be noted that the 3WJ decays show increased values of A_3 and the A_3/A_2 ratio, compared with duplex, in all cases, but in most cases, the values of A_3 and A_3/A_2 ratio for 3WJ is less than for ssDNA, as shown in Figure 3.17. This implies that 2AP in all positions in the 3WJ sees a weakening of base paring compared with the duplex. In 3WJ6 the A_3/A_2 ratio most closely approaches that of the corresponding ssDNA, suggesting significant unzipping.

3WJ	A-Factor				Lifetimes/ns			
	A_1	A_2	A_3	A_4	τ_1	τ_2	τ_3	τ_4
3WJ1	0.66	0.18	0.13	0.04	0.07	0.60	2.80	10.3
3WJ2	0.79	0.11	0.07	0.02	0.05	0.47	2.39	10.1
3WJ3	0.75	0.14	0.09	0.02	0.08	0.45	2.22	8.1
3WJ4	0.71	0.16	0.10	0.02	0.06	0.50	2.36	9.5
3WJ5	0.61	0.30	0.08	0.02	0.08	0.38	2.13	7.9
3WJ6	0.64	0.17	0.15	0.04	0.07	0.59	2.85	9.1
3WJ7	0.68	0.18	0.11	0.03	0.08	0.57	2.41	9.3

Table 3.5—Results derived from global analysis of fluorescence decay curves for 2AP-labelled 3WJ in buffered aqueous solution (reported A-factors are the average value for 2AP emission at 370, 380 and 390 nm).

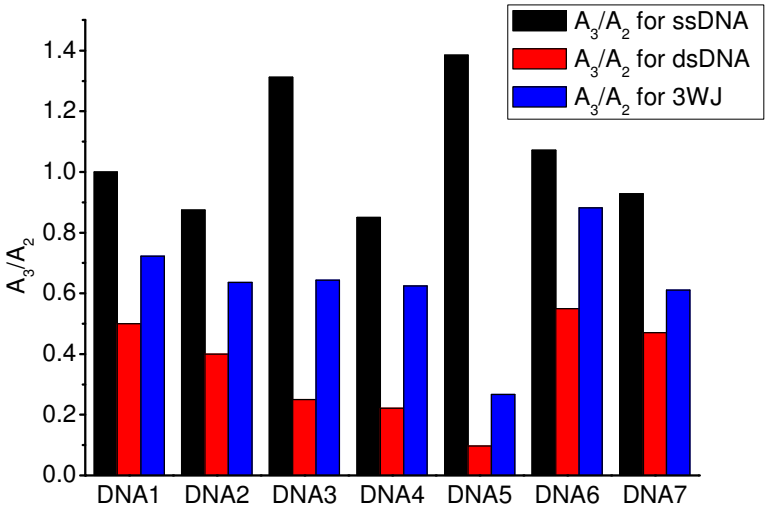
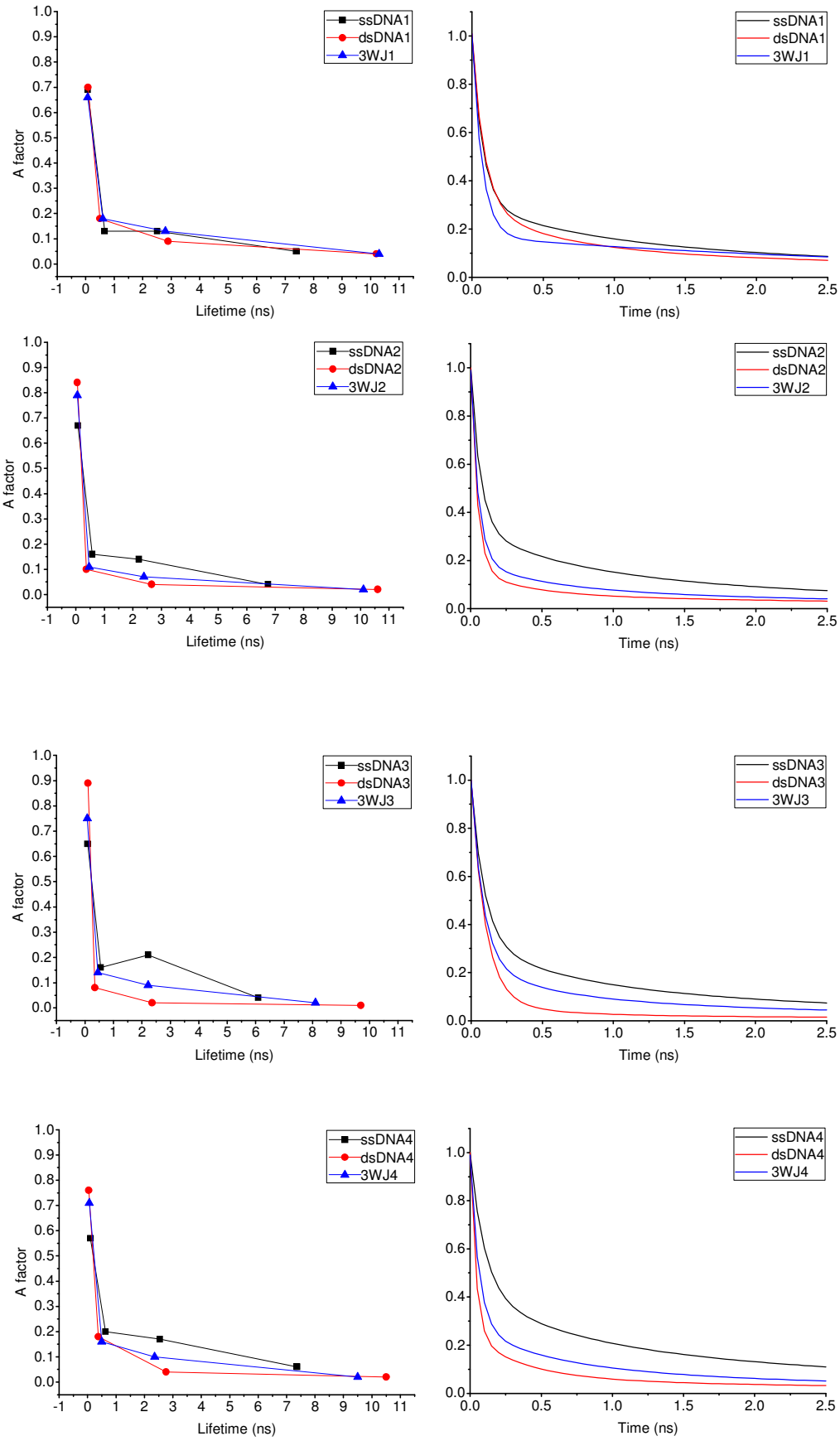


Figure 3.17—The ratios of A_3 to A_2 for 2AP in the various positions in the 3WJ and the corresponding ssDNA and dsDNA samples.

The extent of unzipping at the different 2AP positions can not be judged simply on the basis of the A_3/A_2 ratio. It is necessary to determine, in each case, whether the decay parameters overall are closer to those of the corresponding dsDNA or ssDNA. To assist this comparison, Figure 3.18 presents plots of A *vs.* τ for ssDNA, dsDNA and 3WJ samples, together with the early portions of the fitted decay curves.


ssDNA1
dsDNA1
3WJ1
A factor
Lifetime (ns)
Time (ns)
ssDNA2
dsDNA2
3WJ2
A factor
Lifetime (ns)
Time (ns)
ssDNA3
dsDNA3
3WJ3
A factor
Lifetime (ns)
Time (ns)
ssDNA4
dsDNA4
3WJ4
A factor
Lifetime (ns)
Time (ns)

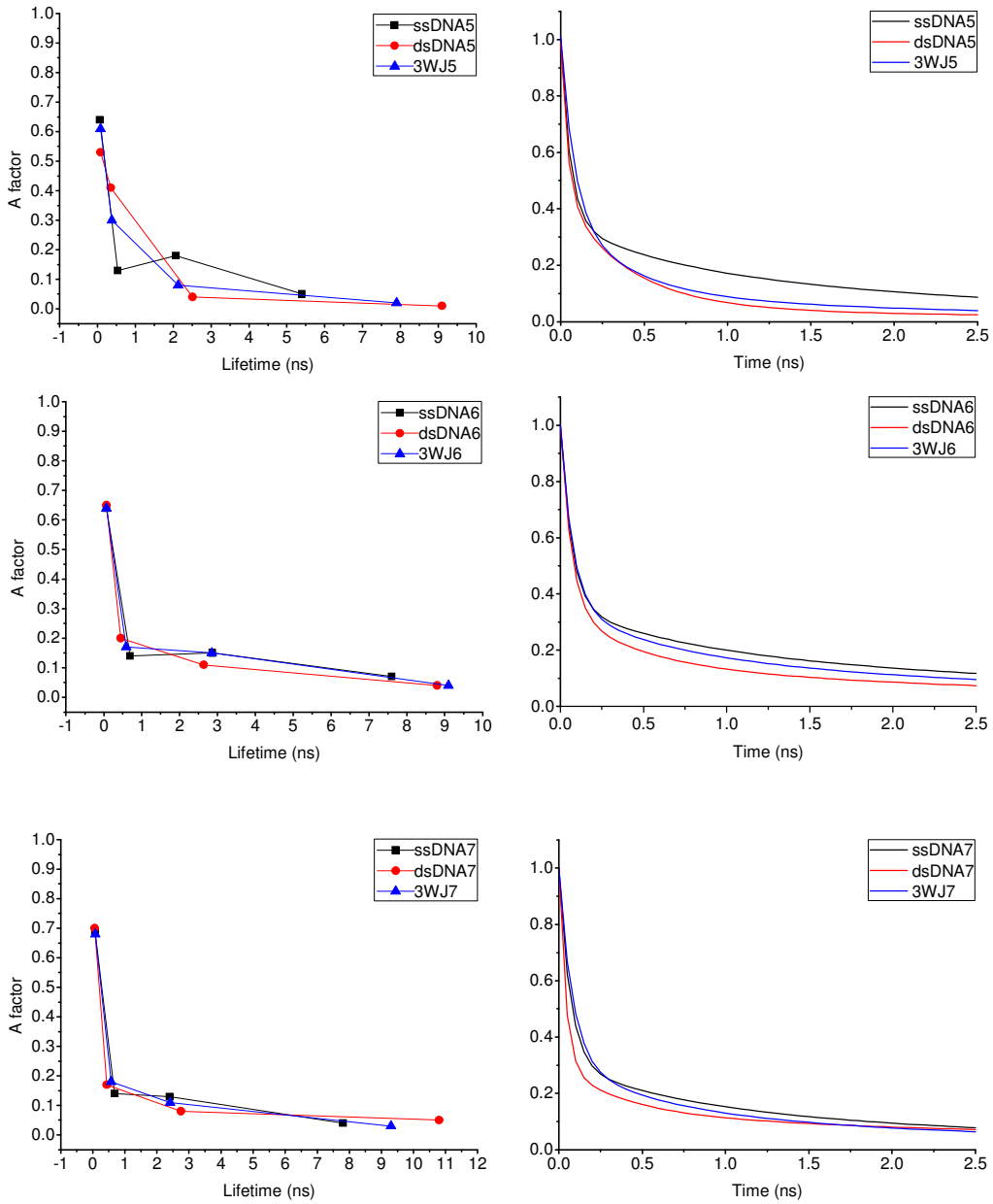


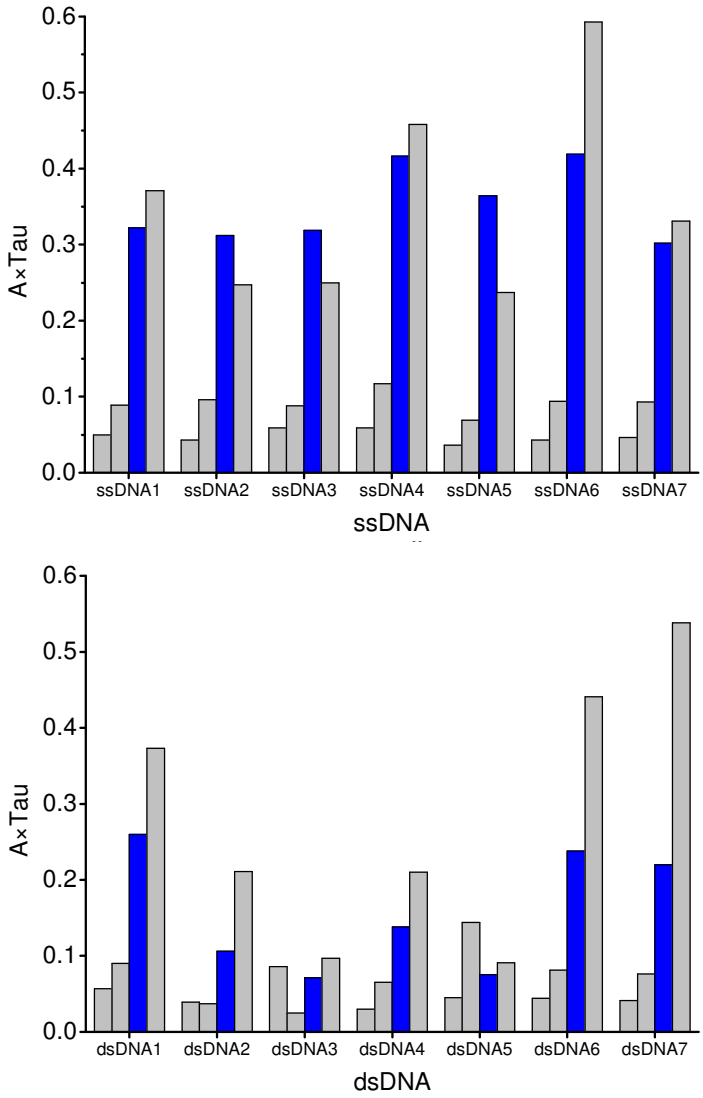
Figure 3.18—Plots of A-factor *vs.* lifetime (left panel) and the early part of the fitted fluorescent lifetime decays (right panel) for ssDNA, dsDNA and 3WJ samples.

When scrutinising the decay parameter of 2AP in 3WJ compared with ssDNA and dsDNA, we will begin with 2AP in the arms of the 3WJ, where we would expect to see duplex-like structure. For 2AP in DNA4, which shows a large change in parameters between ssDNA and dsDNA; the 3WJ clearly resembles dsDNA in terms of the decay pattern. Effects of quenching of Gs in the opposite strand for 3WJ4 and dsDNA4 result in low value of τ_1 and high value of A_1. However, a notable increase

in A_3 relative to dsDNA implies weakening of local base pairing. For 2AP at position 5, 2AP is located in TATA box, which shows very marked difference between dsDNA and ssDNA in lifetime parameters. The 3WJ parameters are clearly characteristic of dsDNA structure in this case, with the unusual high value of A_2 retained (A_2 is 0.30 and 0.13 for 3WJ5 and ssDNA4, respectively). For 2AP at position 1, the change in parameters between dsDNA and ssDNA is quite small, but the 3WJ more closely resemble dsDNA, which is largely revealed by the fitted decay curves in Figure 3.18. For 2AP at position 7, which also shows relatively small change in parameters between dsDNA and ssDNA, the pattern of parameters for 3WJ (A *vs*. τ plot) more closely resembles dsDNA and A_2 is significantly larger than A_3, also consistent with a dsDNA-like conformation. However, it is clear that the structure deviates from a perfect duplex.

Next, the author set to examine 2AP at positions near the branch point. The total length of the DNA strand composing the 3WJ is 50 mer, therefore the 25^{th}, 26^{th} base (from 5' to 3') can be deemed as exact the branch point. 2AP in 3WJ2 is the 23^{rd} base, two bases away from the branch point), 2AP in 3WJ6 (24^{th}) is one base away from the branch point and 2AP in 3WJ3 (25^{th}) is right in the branch point. It is clear from Figure 3.18 that 3WJ2 resembles dsDNA2. As noted above, 2AP in dsDNA2 is highly stacked due to its location in a C:G rich region. Though position 2 is close to the branch point, the well-stacked duplex structure remained in the 3WJ. This reflects the influence of sequence context on the local conformation. For 2AP in position 3, 2AP exactly located at branch point as well as being in a C:G-rich sequence context (giving a very high A_1 in dsDNA3). Distortion of the duplex structure in 3WJ3 is clearly apparent as evidenced by a decrease in A_1, and increase in A_2 and A_3, compared to dsDNA3. It can be seen in Figure 3.18 that the decay parameters for 3WJ3 lie between those of dsDNA3 and ssDNA3, showing a tendency towards ssDNA. Besides, in agreement with the analysis for decay parameters, the decay curves also show that 3WJ3 lies in between ssDNA3 and dsDNA3, but closer to ssDNA. We come finally to 3WJ6, in which 2AP is one base away from the branch point and which shows decay parameters that resemble ssDNA: the value of A_3 approaches that of A_2; τ_3 is equal to the value in ssDNA, τ_2 is also closer to that of

ssDNA than dsDNA. As shown in Figure 3.18, the decay curve of 3WJ6 is closer to this of ssDNA6. All these observations imply an appreciable local unzipping at this position. Taken overall, it can be observed that for each position, no matter if it is in the arms or near the branch point, there exists some distortion, but the extent for the latter is more marked than for the former.



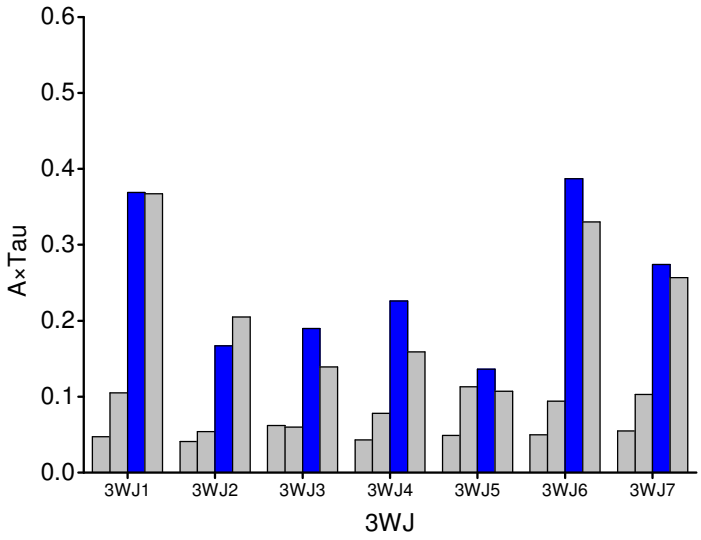


Figure 3.19—Graphical representation of decay parameters for ssDNA, dsDNA and 3WJ samples. The four A-factor×lifetime components are listed from left to right for different samples. The 3rd components are highlighted in blue.

The author believes that the combination of the first two approaches (A-factor *vs*. lifetime plots and decay curves) can give a persuasive demonstration for looking at unzipping. However, a third approach of presenting data and showing unzipping effect can be introduced as well. As shown in Figure 3.19, for DNA3, component 2 of 3WJ3 shows much increase compared with duplex, and component 3 remains smaller in ssDNA, suggesting unzipping. For DNA6, component 3 stands out for 3WJ6, almost as high as ssDNA6, and component 3 and 4 for 3WJ6 are much greater than component 1 and 2. All these mean that 2AP in 3WJ6 is in a very flexible conformation which is alike ssDNA, indicative of unzipping in 3WJ6. The 3rd component for both 3WJ4, 3WJ5 also stands out but much smaller than ssDNA, indicating some base disturbances.

3.4 Summary and Conclusions

Overall, fluorescence lifetimes of 2AP for ssDNA and dsDNA samples are both manifested by four discrete lifetime components, showing comparable conformational heterogeneity. The highly quenched/stacked base conformation is retained in ssDNA and accounts for the majority of the population (more than 50%)

for both ssDNA and dsDNA. For dsDNA, there is a general tendency for the A-factor to decrease with increasing lifetime, which is highly in line with the findings of previous studies.

In terms of the difference between ssDNA and the corresponding dsDNA, the most constant and characteristic change is in the relative populations of τ_2 and τ_3. For dsDNA, A_2 is significantly greater than A_3, whereas for ssDNA these two values are similar. It is inferred that for ssDNA, population is transferred from the 1st and 2nd lifetime components to the 3rd lifetime component, namely from a relatively well based-stacked conformations to a poorly base-stacked conformations, as a consequence of removing the complementary strand. The second feature observed is the considerable shortening of τ_4 for ssDNA compared with dsDNA. For ssDNA, the mobility of the neighbouring base(s) allows them to move in a concerted fashion with the adjacent 2AP base and a more hydrophobic microenvironment is thus maintained.

The local base dynamics at different positions in a 3WJ was successfully analysed by comparison of the 2AP decay parameters with those of corresponding ssDNA and dsDNA samples, using the pattern of the A-factor *vs*. τ plots and the decay curves as graphical indicators. Our experiments, in the first place, demonstrate that ultra-fast time-resolved fluorescence measurements of 2AP-labelled oligonucleotides is a practical approach to investigate the base dynamics of branched nucleic acids, such as 3WJ. Although, there are previous reports of application of this technique to nucleic acid dynamics,[44-47] to our knowledge, this is the first time for it to be adopted for the study of 3WJ. These experiments confirm the proposal of our collaborators, based on SM-FRET and molecular dynamics simulations, those bases at the branch point of the perfectly base-paired DNA 3WJ tend to unzip, in spite of the full complementarity of the sequence. The 2AP decay parameters also reveal that bases in the arms of the 3WJ junction, more than 10 base pairs away from the branch point, are base paired less perfectly than dsDNA. At all of the positions studied, 2AP in the arms of 3WJ shows local conformational distortion from the perfect duplex structure to some extent (but less than for 2AP bases near the branch point). These results

support previous findings of a very dynamic structure in both branch point and helical arms of 3WJ. DNA cyclisation experiments suggested that the 3WJ is a very flexible structure.[48] The branches of 3WJ were postulated to have variable angles of 60-90° between arms, indicative of dynamic conformation.[49] It was proposed that there might exist local unfolding/plasticity in the branch point region, in which at least one base pair is unzipped transiently.[50]

If constructing a model of the 3WJ from fully base-paired B-DNA, simple geometrical arguments suggest that it will be necessary to distort the first helix in order to introduce the third single-stranded DNA arm, generating disrupted base stacking. This changed base stacking would not be identical to B-form DNA, which would account for the base 'dangling' in 3WJ.

Similar 'dangling bases' have been reported at replication forks and primer-template junctions previously.[51] The role of DNA breathing has been discussed in the context of recognition, binding or activity of DNA-processing enzymes. Jose *et al.*[52] speculated that this unusual more 'open state' of bases perhaps serves as a 'functional position' to enable the specific binding and targeting of DNA enzymes. Also, Ramreddy *et al.*[44] conjectured that DNA base dynamics can be used as a cue for searching system of certain enzymes to recognise a specific part of genome DNA.

3.5 References

1. D. M. J. Lilley and Editor, *DNA-Protein: Structural Interactions*, IRL, 1995.
2. F. Jensch and B. Kemper, *EMBO J.*, 1986, 5, 181-189.
3. R. C. Heller and K. J. Marians, *Nat. Rev. Mol. Cell Biol.*, 2006, 7, 932-943.
4. J. Atkinson and P. McGlynn, *Nucleic Acids Res.*, 2009, 37, 3475-3492.
5. S. M. Hecht and Editor, *Bioorganic Chemistry: Nucleic Acids*, Oxford Univ Press, 1996.
6. D. M. J. Lilley, *Q. Rev. Biophys.*, 2000, 33, 109-159.
7. T. Minagawa, A. Murakami, Y. Ryo and H. Yamagishi, *Virology*, 1983, 126, 183-193.

8. N. N. Degtyareva, C. A. Barber, M. J. Reddish and J. T. Petty, *Biochemistry*, 2011, 50, 458-465.
9. R. R. Sinden, V. N. Potaman, E. A. Oussatcheva, C. E. Pearson, Y. L. Lyubchenko and L. S. Shlyakhtenko, *J. Biosci.* , 2002, 27, 53-65.
10. D. Duzdevich, J. Li, J. Whang, H. Takahashi, K. Takeyasu, D. T. F. Dryden, A. J. Morton and J. M. Edwardson, *PLoS ONE*, 2011, 6, e17119.
11. K. R. Birikh, P. A. Heaton and F. Eckstein, *Eur. J. Biochem.*, 1997, 245, 1-16.
12. R. M. Jimenez, E. Delwart and A. Luptak, *J. Biol. Chem.*, 2011, 286, 7737-7743.
13. L. Huang, A. Serganov and D. J. Patel, *Mol. Cell*, 2010, 40, 774-786.
14. C. Lin, Y. Liu and H. Yan, *Biochemistry*, 2009, 48, 1663-1674.
15. A. L. Benvin, Y. Creeger, G. W. Fisher, B. Ballou, A. S. Waggoner and B. A. Armitage, *J. Am. Chem. Soc.*, 2007, 129, 2025-2034.
16. H. Ozhalici-Unal and B. A. Armitage, *ACS Nano*, 2009, 3, 425-433.
17. N. C. Seeman, *Nano Lett.*, 2010, 10, 1971-1978.
18. F. A. Aldaye, A. L. Palmer and H. F. Sleiman, *Science*, 2008, 321, 1795-1799.
19. S. Xiao, *Shengming De Huaxue*, 2011, 31, 173-183.
20. A. Oleksy, A. Oleksi, A. G. Blanco, R. Boer, I. Uson, J. Aymami, A. Rodger, M. J. Hannon and M. Coll, *Angew. Chem., Int. Ed.*, 2006, 45, 1227-1231.
21. I. T. Seemann, V. Singh, M. Azarkh, M. Drescher and J. S. Hartig, *J. Am. Chem. Soc.*, 2011, 133, 4706-4709.
22. D. R. Boer, J. M. C. A. Kerckhoffs, Y. Parajo, M. Pascu, I. Uson, P. Lincoln, M. J. Hannon and M. Coll, *Angew. Chem., Int. Ed.*, 2010, 49, 2336-2339.
23. D. R. Duckett and D. M. J. Lilley, *EMBO J.*, 1990, 9, 1659-1664.
24. J. B. Welch, D. R. Duckett and D. M. J. Lilley, *Nucleic Acids Res.*, 1993, 21, 4548-4555.
25. J. Ouellet, S. Melcher, A. Iqbal, Y. Ding and D. M. J. Lilley, *RNA*, 2010, 16, 1597-1609.
26. M. Yang and D. P. Millar, *Biochemistry*, 1996, 35, 7959-7967.
27. F. Stuehmeier, J. B. Welch, A. I. H. Murchie, D. M. J. Lilley and R. M. Clegg, *Biochemistry*, 1997, 36, 13530-13538.
28. V. Thiviyanathan, B. A. Luxon, N. B. Leontis, N. Illangasekare, D. G. Donne

and D. G. Gorenstein, *J. Biomol. NMR*, 1999, 14, 209-221.

29. B. Wu, F. Girard, B. B. Van, J. Schleucher, M. Tessari and S. Wijmenga, *Nucleic Acids Res.*, 2004, 32, 3228-3239.
30. E. A. Oussatcheva, L. S. Shlyakhtenko, R. Glass, R. R. Sinden, Y. L. Lyubchenko and V. N. Potaman, *J. Mol. Biol.*, 1999, 292, 75-86.
31. L. S. Shlyakhtenko, V. N. Potaman, R. R. Sinden, A. A. Gall and Y. L. Lyubchenko, *Nucleic Acids Res.*, 2000, 28, 3472-3477.
32. E. Sisamakis, A. Valeri, S. Kalinin, P. J. Rothwell and C. A. M. Seidel, *Methods Enzymol.*, 2010, 475, 455-514.
33. A. T. Brunger, P. Strop, M. Vrljic, S. Chu and K. R. Weninger, *J. Struct. Biol.*, 2011, 173, 497-505.
34. S. Weiss, *Science*, 1999, 283, 1676-1683.
35. R. Roy, S. Hohng and T. Ha, *Nat. Methods*, 2008, 5, 507-516.
36. S. A. McKinney, A.-C. Declais, D. M. J. Lilley and T. Ha, *Nat. Struct. Biol.*, 2003, 10, 93-97.
37. T. Sabir, G. F. Schrodr, A. Toulmin, P. McGlynn and S. W. Magennis, *J. Am. Chem. Soc.*, 2011, 133, 1188-1191.
38. T. Sabir, A. Toulmin, L. Ma, A. C. Jones, P. McGlynn, G. F. Schroder and S. W. Magennis, *J. Am. Chem. Soc.*, 2012, 134, 6280-6285.
39. E. Y. M. Bonnist, K. Liebert, D. T. F. Dryden, A. Jeltsch and A. C. Jones, *Biophys. Chem.*, 2012, 160, 28-34.
40. R. K. Neely, D. Daujotyte, S. Grazulis, S. W. Magennis, D. T. F. Dryden, S. Klimasauskas and A. C. Jones, *Nucleic Acids Res.*, 2005, 33, 6953-6960.
41. R. K. Neely and A. C. Jones, *J. Am. Chem. Soc.*, 2006, 128, 15952-15953.
42. R. K. Neely, G. Tamulaitis, K. Chen, M. Kubala, V. Siksnys and A. C. Jones, *Nucleic Acids Res.*, 2009, 37, 6859-6870.
43. P. Rai, T. D. Cole, E. Thompson, D. P. Millar and S. Linn, *Nucleic Acids Res.*, 2003, 31, 2323-2332.
44. T. Ramreddy, Rao and G. Krishnamoorthy, *J. Phys. Chem. B*, 2007, 111, 5757-5766.
45. E. L. Rachofsky, E. Seibert, J. T. Stivers, R. Osman and J. B. A. Ross, *Biochemistry*, 2001, 40, 957-967.

46. L. J. Reha-Krantz, C. Hariharan, U. Subuddhi, S. Xia, C. Zhao, J. Beckman, T. Christian and W. Konigsberg, *Biochemistry*, 2011, 50, 10136-10149.
47. R. A. Hochstrasser, T. E. Carver, L. C. Sowers and D. P. Millar, *Biochemistry*, 1994, 33, 11971-11979.
48. R. I. Ma, N. R. Kallenbach, R. D. Sheardy, M. L. Petrillo and N. C. Seeman, *Nucleic Acids Res.*, 1986, 14, 9745-9753.
49. L. S. Shlyakhtenko, D. Rekesh, S. M. Lindsay, I. Kutyavin, E. Appella, R. E. Harrington and Y. L. Lyubchenko, *J. Biomol. Struct. Dyn.*, 1994, 11, 1175-1189.
50. N. C. Seeman and N. R. Kallenbach, *Annu. Rev. Biophys. Biomol. Struct.*, 1994, 23, 53-86.
51. N. C. Seeman and N. R. Kallenbach, *Annu. Rev. Biophys. Biomol. Struct.*, 1994, 23, 53-86.
52. D. Jose, K. Datta, N. P. Johnson and P. H. von Hippel, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, 106, 4231-4236.

Chapter 4---Investigation of Double-Nucleotide Unzipping Mechanism for Human Flap Endonuclease

4.1 Introduction

4.1.1 FEN and its proposed double-nucleotide unzipping mechanism

Flap endonuclease 1 (FEN1) proteins, present in all kingdoms of life, are able to catalyse the sequence-independent hydrolysis of the bifurcated nucleic acid intermediates formed during DNA replication and repair, which is essential for maintaining genomic stability.[1] Critical to replication efficiency, the FEN1-cleaved product from a double flap can be ligated without further processing, in spite of the exceptional stability of RNA and DNA phosphodiester bonds against attack by water or hydroxide.[2,3] FEN1 enhances the hydrolysis rate of targeted phosphodiester bonds up to ~10^{17}-fold.

FEN1 removes 5' flaps in a structure-specific manner using divalent metal ion-dependent catalysed phosphodiesterase activity, as reported by Tomlinson *et al*.[4] The importance of FEN can be evidenced by the fact that *fen1* deletion is embryonically lethal in mice. Consistent with its crucial role in DNA replication, human FEN1 (hFEN) is highly expressed in all proliferative tissues. Nevertheless, many cancers show even higher levels of FEN1 expression, and in several cases this is correlated with tumour aggressiveness.[5] Thus, FEN1 specific inhibitors potentially could be used as chemotherapeutic reagents.[6]

Tsutakawa *et al.*[7] determined the crystal structure of hFEN1 in complex with its cognate DNA, before (WT hFEN1:substrate DNA, PDB: 3Q8L,) and after cleavage (WT hFEN1:product DNA, PDB: 3Q8K) in the presence of Samarium (III). The structure of D181A hFEN1:substrate DNA was also obtained (PDB: 3Q8M). Figure 4.1 shows the structure of WT hFEN1 in complex with product DNA.

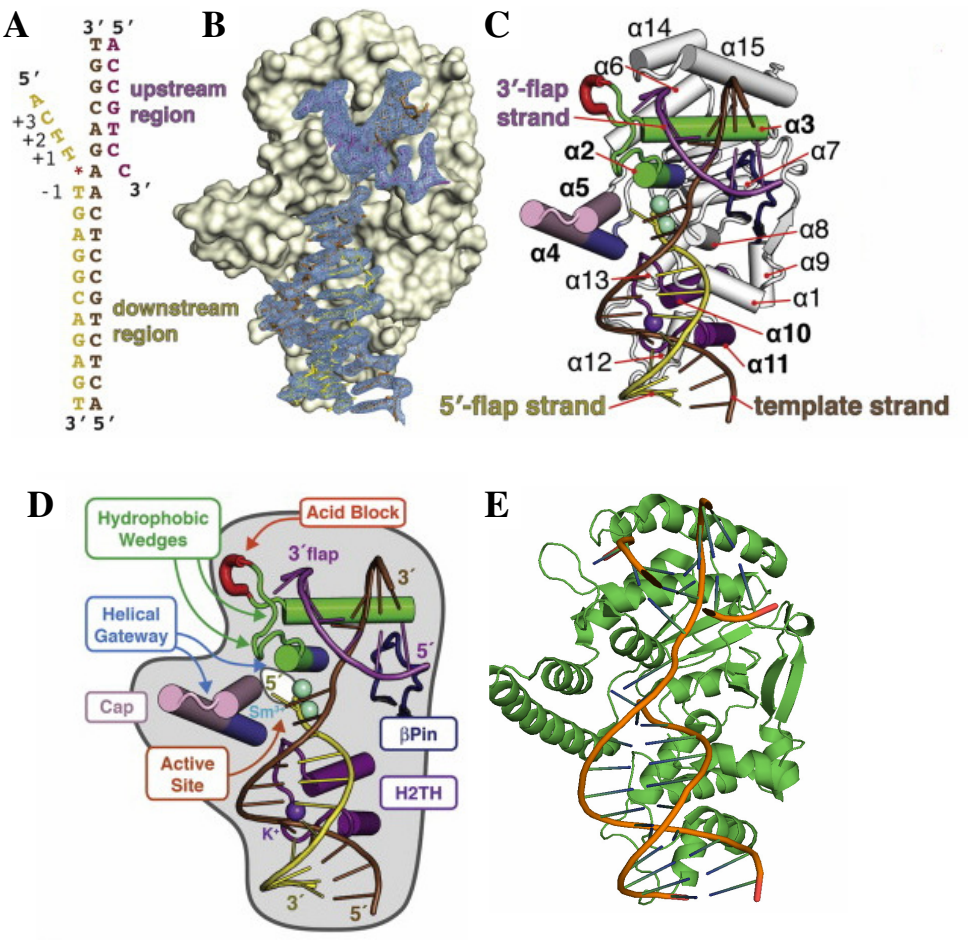


Figure 4.1—Structural analysis of hFEN1:product DNA complex in the presence of Sm^{3+} (reprinted from Tsutakawa *et al.*[7]). (A) The sequence of substrate DNA used. The 4 bp nucleotide flap substrate is numbered relative to the scissile phosphate. (B) hFEN1's surface electron density shows clear density for bent DNA bound by hFEN1. (C) Structural elements of hFEN1 with respect to its functions. Key elements include the hydrophobic wedge that breaks DNA path and forms the 3' flap binding site (green), acid block that inhibits longer 3' flaps (red), helical gateway that permits only ssDNA and forms the active site (dark blue), helical cap that imposes the preference for 5' termini (pink), and H2TH that binds to dsDNA upstream from 5'flap (purple). Sm^{3+} (cyan) and K^{+} (purple) ions are shown as spheres. (D) Crystal structure of hFEN1:product DNA is shown in cartoon mode (PDB: 3Q8K)

By comparing the crystal structures of the above-mentioned enzyme-DNA complexes, Tsutakawa *et al.*[7] revealed a double-nucleotide unzipping, flanking the scissile phosphate, which hFEN1 adopts for controlling precise flap incision by the two-metal-ion active site. In the WT and D181A substrate structures, the two bases, +1 and -1, on either side of the scissile phosphate are base paired to the template strand, unlike the product complex in which the -1 base is clearly unzipped, as shown in Figure 4.2A-Figure 4.2D. Tyr40 (Y40) is stacked with the +1 base in the hFEN1:substrate DNA complexes, while it rotates 25° to stack with the -1 flap base in the hFEN1:product DNA complex. The template strand is displaced only slightly between the product and substrate complexes. Since the -1 base and two terminal bases in template strand are unzipped in hFEN1: product DNA complex as shown in Figure 4.2D, two bases (+1 and -1 base) of substrate must be unzipped to position the scissile phosphate between the divalent metal ions when fitting in the active site of hFEN1. These results indicate an important specificity step in the mechanism in between binding and catalysis. They found that in the substrate complexes, the region bound by the H2TH and K^+ ion is in line with B-DNA parameters, whereas the dsDNA near the active site deviates from canonical B-DNA geometry. Actually, the +1 base pair (A:T) in the substrate forms only one of two possible H-bonds due to a pronounced base pair opening of ~20° and stagger of 0.61Å toward the major groove. Taken together, these disparities suggest that hFEN1 induces deformations in the downstream dsDNA base pairs closest to the active site to facilitate the requisite unzipping for providing specific incision one base away from the dsDNA region.

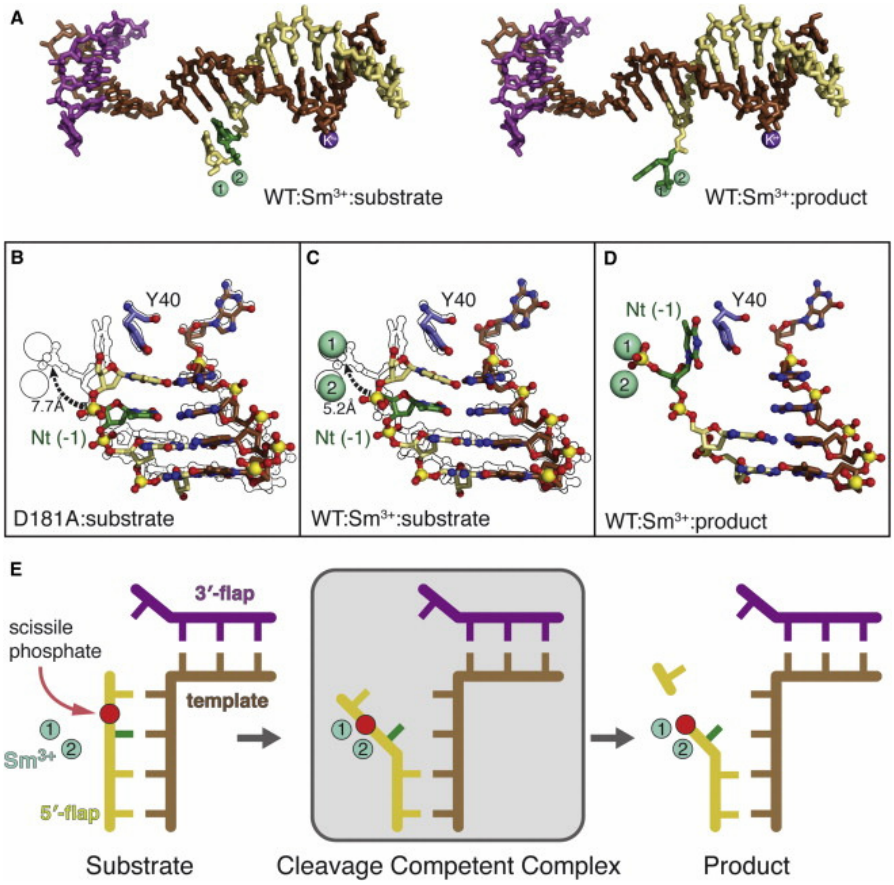


Figure 4.2—Comparison of substrate and product complex DNA structure in the FEN-DNA complex reveals a double-nucleotide unzipping mechanism for achieving scissile phosphate positioning (reprinted from Tsutakawa *et al.*[7]). (A) In WT hFEN1:substrate DNA complex, the +1 and -1 base relative to scissile phosphodiester bond are paired in the substrate but the -1 base in the hFEN1:product DNA complex is unzipped with Sm^{3+} ions. (B) Close-up of the DNA and Tyr40 from the D181A:substrate complex showing the base-pairing of the +1 and -1 position and Tyr40 stacking with the +1 nt. The overlay with the product structure outline highlights the 7.7 Å movement of the scissile phosphate of the -1 base (green) needed for catalysis. (C) as in (B), the +1 and -1 base are base-paired and -1 base is stacked with Tyr40 in the WT hFEN1:Sm^{3+}:substrate. The scissile phosphate of the -1 base moves 5.2 Å into the active site. (D) Close-up view of the DNA and Tyr40 from the WT:Sm^{3+}:product DNA complex. Unlike the substrate complexes, the +1 base has been cleaved off, the -1 base is unzipped, and Tyr40 stacks with the -1 base. (E) Proposed model of double-nucleotide unzipping to move substrate into an active site for precise incision.

The findings reveal that hFEN1 adopts an extraordinarily sophisticated mechanism to achieve structure-specific, sequence-independent recognition and catalysis on its substrates. A 'bind-push-pull-unzipping' mechanism was proposed, based on the structures and biochemical study, as illustrated in Figure 4.3. To be detailed, first, FEN1 recognizes the 5' flap DNA by its ability to form a sharp ~100° bend with dsDNA on either side. Only dsDNA with a flap or break can bend at a single phosphodiester link to that degree. Second, a 3' flap binding pocket encloses a single unzipped base, and this binding of unzipped 3' flap comprises one-eighth of the hFEN1:DNA binding surface. This can explain the observation of by Finger *et al.*[8] where they found that hFEN1 shows an 33-fold increase in nuclease activity for 5' flap substrates with an unzipped 3' flap, compared to substrates without a 3' flap.[8] Biologically, this ensures that the FEN1 product is suitable for ligation. Mechanistically, the 3' flap binding and the DNA bend establish the register of the incision position by pushing the 5'-flap through and under the disordered helical gateway and cap. Third, hFEN1 requires the 5' flap to pass under the cap to enter the helical gateway to reach into the active site. Only DNA with ssDNA portion can do so. Fourth, the substrate-induced, disorder-to-order transition in the helical cap takes place and hFEN1 binds the template strand at the DNA bend and at the K^+:H2TH, thereby pulling the strand complementary to it, the 5' flap strand, toward the active site. Biologically, this template-based positioning of the 5' flap prevents inadvertent incision of ssDNA. Last, the entrance to the hFEN1 active site is guarded on two sides by the α2-α4 helical gateway, in order to only allow the passage of ssDNA rather than dsDNA into the active site. The scissile phosphodiester is only positioned at the active site with divalent metals and the double-nucleotide (+1 and -1) unzipping of the 5' flap ensures the access to the two-metal-ion active site to achieve the specificity.

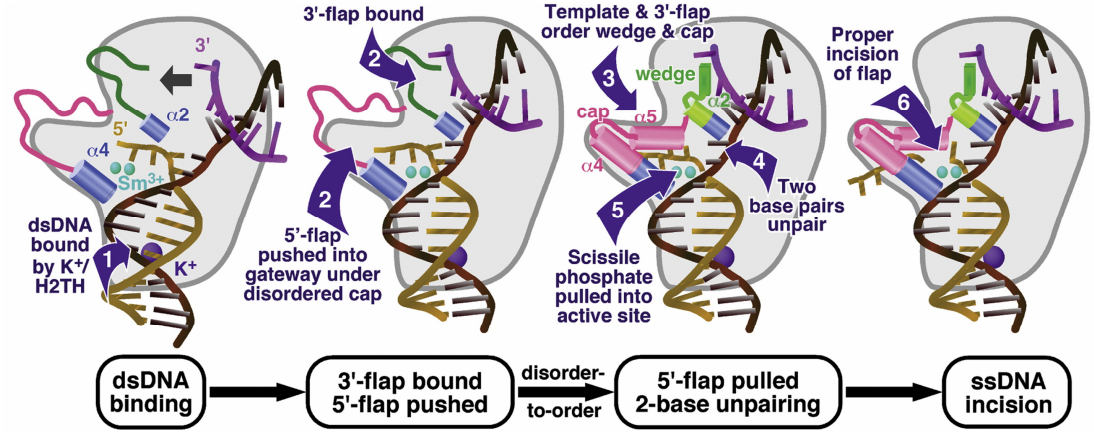


Figure 4.3—Illustration of the mechanism of hFEN1 binding, subsequent bends on substrate dsDNA junctions and undergoes disorder-to-order transitions for 5' flap recognition (reprinted from Tsutakawa *et al.*[7]). hFEN1 binds primarily to downstream dsDNA by the K^+:H2TH, allowing FEN1 to search for target DNA structure. The second binding site selects for dsDNA structures that can sharply bend ~100°. Coincident with DNA template strand binding, the 5' flap is directed under the disordered cap domain and through the helical gateway selecting for ssDNA nearing the active site. Collectively, the 3' flap site biding, disorder to order conformational change in helical cap, and double-nucleotide unzipping promote correct positioning of the scissile bond and rapid two-metal-ion catalysed incision.

As Grasby *et al.*[9] proposed, unzipping together with gating is a unified mechanism for flap endonuclease and FEN superfamily nucleases, including exonuclease 1, DNA repair protein XPG, endonuclease GEN1 and the 5'-3' exoribonucleases, to achieve sequence-independent substrate recognition, in spite of substrate diversity. This is illustrated in Figure 4.4. The entrance to the active site is restricted by the gateway formed by two α helices that only allow single-stranded nucleic acid to enter. The terminal bases of the 5'-strand must enter the helical gateway for incision.

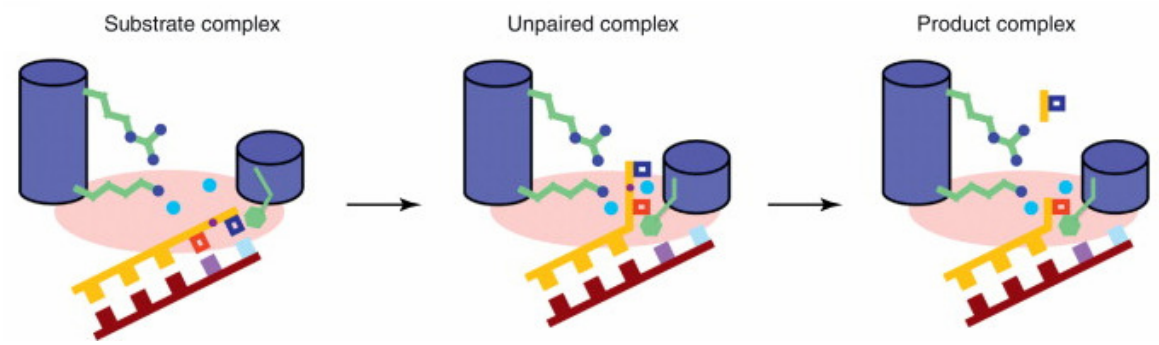


Figure 4.4—Illustration of double-nucleotide unzipping mechanism required for substrates for entering the active site of enzymes of the 5'-nuclease superfamily (reprinted from Grasby *et al.*[9]). A double-nucleotide unzipping and gateway mechanism for substrate selection FEN superfamily enzymes is depicted. The gateway (blue cylinder), the highly conserved lysine, arginine and stacking residues (coloured in green) and the active site carboxylates (pink oval) are depicted. Initially, the terminal two nucleotides in substrate are base-paired (blue with light blue, red with violet), in which the scissile phosphodiester bond does not enter the active site and contact active site metals (cyan) in the substrate complex. Double-nucleotide unzipping allows the scissile phosphodiester (brown dot) to enter the active site and contact the active site metals (unzipped complex), resulting in reaction and the formation of product.

4.1.2 Aims of this chapter

2AP bases were used to substitute the nucleobases at positions -1 and +1 with respect to the scissile phosphodiester bond on the primer strand, to serve as a structural probe for investigating the stacking interactions of each nucleobase upon enzyme manipulation in the enzyme-substrate complexes. By comparison of the fluorescence lifetime of 2AP in -1 and +1 positions in free duplex with those in complex with either WT hFEN1 or the three mutants (Y40A, R100A, K93A), an insight should be obtained into the local changes in DNA conformation induced by the enzyme in the transition from enzyme-substrate to enzyme-product complex. The reason for studying these three mutants mentioned above can be found as follows. It has been reported that Y40A reduced the reaction rate 20-fold, supporting a significant role in activity likely through substrate positioning.[7] Lys93 and Arg100 interact with the product 5'-phosphate monoester in our structures, probably play an important role in stabilising the unpaired conformation of DNA substrates and act as electrostatic catalysts during hydrolysis.[7] K93A and R100A dramatically reduced the rate of the

FEN1 catalysed reaction by >400-fold.[7] Additionally, Lys93, Arg100 and stacking residues (in FEN case, it is Y40) are hypothesized to play an important role in double-nucleotide unzipping in the 5'-nuclease superfamily (as shown in Figure 4.4).[9] By relating the fluorescence lifetime signals to the predicted changes, the proposed mechanism of double-nucleotide unzipping could be tested.

4.2 Experimental

4.2.1 DNA sequences, enzymes and buffer

DNA was obtained from Dr. Jane Grasby (School of Chemistry, University of Sheffield, UK), the sequences are shown below. 2AP-1 is the sequence containing a 2AP base in the +1 position, while 2AP-1 is the sequence containing a 2AP base in the -1 position. 2AP-9 contains one 2AP base near the end of the duplex, remote from the scissile phosphate.

Figure 4.5 shows the three alternative positions of 2AP in the duplex formed between the 2AP-labelled primer strand and the template strand. Proteins used in this study were WT hFEN, and three mutants Y40A, K93A, R100A, which were also provided by Dr Jane Grasby.

2AP+1 primer strand:
5'-TTTTT(2AP)AGAGGCAGAGTG

2AP-1 primer strand:
5'-TTTTTA(2AP)GAGGCAGAGTG

2AP-9 control primer strand:
5'-TTTTTAAGAGGCAG(2AP)GTG

Template strand:
5'-CACTCTGCCTCTTGACGGT[GAA]ACCGTCC-3' (Highlighted terminal C in orange is one nucleotide 3' flap)

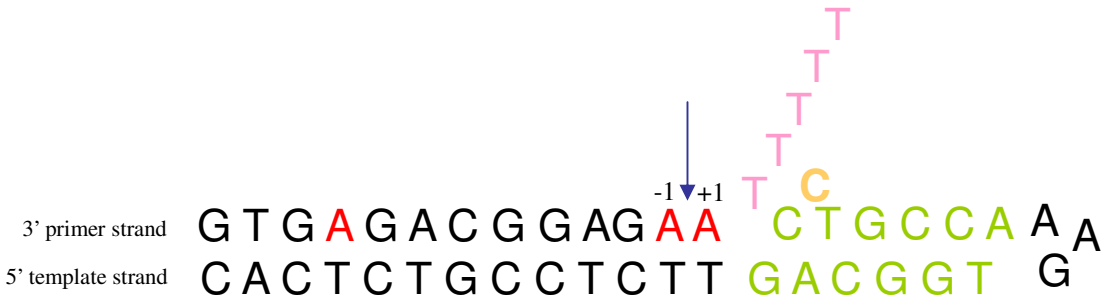


Figure 4.5—The structure of the duplex with 2AP substitution. A coloured in red represents 2AP base, +1 to the right and -1 to the left of the scissile phosphodiester bond. The five-base flap on the 5' end is shown in pink and the one-base flap on 3' end is shown in orange. The upstream duplex is in green (except hairpin turn). The scissile phosphate is indicated by blue arrow.

For free DNA duplex measurements, each DNA duplex was buffered with 50 mM HEPES, 100 mM KCl. For DNA duplex+enzyme, 10 μM duplex of each kind was used to react with 10 μM enzyme in the buffer containing 50 mM HEPES, 100 mM KCl, 1 mM DTT 10 mM $CaCl_2$. The mixture was premixed and held at room temperature for ~15 min before the measurement.

Experiments were done for free 2AP-1, 2AP+1, 2AP-9 respectively, 2AP-1 in the presence of each WT hFEN/Y40A/R100A/K93A, 2AP+1 in the presence of each of WT hFEN/Y40A/R100A/K93A and 2AP-9 in the presence of WT hFEN1.

4.2.2 Time-resolved fluorescence measurements

Time-resolved fluorescence was recorded using TCSPC as described in Chapter 2. The fluorescence lifetimes were measured at excitation wavelength 315 nm. The fluorescence lifetime decays were collected at three emission wavelengths: 370 nm, 380 nm and 390 nm respectively. The slits of the monochromator were set to 10 nm. Data processing was carried out as described in Chapter 2, to obtain the lifetimes and their respective fractional amplitudes.

4.3 Results and Discussion

4.3.1 Note on the response of 2AP to duplex unzipping

As shown in the systematic study of 2AP-labelled ssDNA and corresponding dsDNA samples presented in Chapter 3, an unzipping of the duplex is manifested by substantial increase in both A_3 and τ_3. Thus, unzipping can be revealed by plotting the values of the product $A_i \times \tau_i$ for each lifetime parameter,[10] as exemplified in Figure 4.6 for the 2AP+1 dsDNA/ssDNA used in the present study. The decay parameters for the dsDNA and corresponding ssDNA are compared in Table 4.1. In addition to the increase in A_3 and τ_3, it can be seen that the population of highly stacked states (A_1) is decreased from 0.75 in dsDNA to 0.47 in ssDNA, though still substantial in the ssDNA. The degree of quenching is reduced as evidenced by the increase in τ_1 from 0.060 ns in dsDNA to 0.097 ns in ssDNA.

DNA	τ_1/ns (A_1)	τ_2/ns (A_2)	τ_3/ns (A_3)	τ_4/ns (A_4)
2AP+1 duplex	0.060 (0.75)	0.30 (0.17)	1.5 (0.06)	5.7 (0.02)
2AP+1 ssDNA	0.097 (0.47)	0.56 (0.29)	1.8 (0.19)	4.9 (0.05)

Table 4.1—Lifetime parameters of 2AP+1 dsDNA and its corresponding ssDNA.

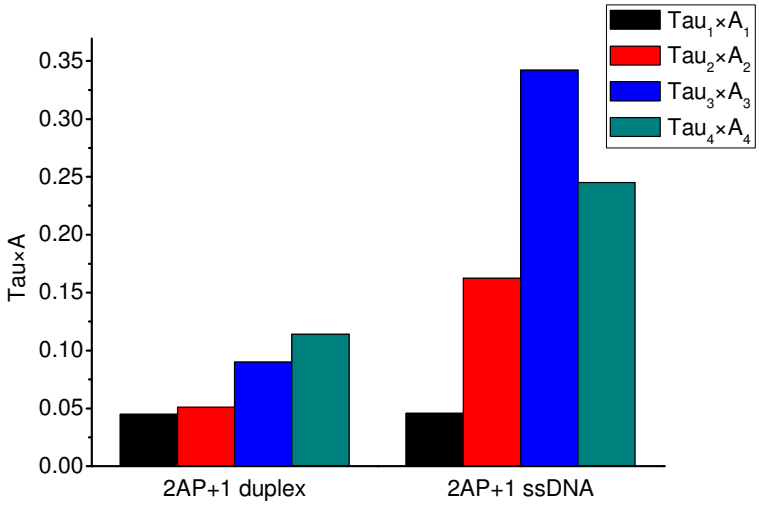


Figure 4.6—Graphical representation of decay parameters for 2AP-labelled duplex 2AP+1 and its corresponding ssDNA.

4.3.2 Lifetime parameters of the free duplexes

All three duplexes show typical lifetimes and fractional amplitudes as reported previously in this thesis and elsewhere.[10-15] In each case, the shortest lifetime component of less than 100 ps accounts for the greatest population, while the longest lifetime represents the smallest population. Invariably, the A-factor declines with the increase of lifetime.

As shown by the values of A_1 for all three duplexes (Table 4.2), at least 75% of the duplex population exists in this highly stacked state. The longest lifetime component is due to 2AP that is largely unstacked and thus free from interbase electron transfer. It is notable that the values of τ_4 for all three duplexes, from 5.7-7.1 ns, are much shorter than that for 2AP in aqueous solution (10.6 ns) and similar to that in less polar solvents, indicating that the 2AP is not fully extrahelical. This unstacked 2AP comprises only a few percent of the total conformational population. The intermediate decay times represent 2AP in an intrahelical, apolar environment, in moderately (τ_2) or poorly (τ_3) stacked states, in which the duplex conformation does not favour rapid interbase electron transfer. τ_1 is evidently longer for 2AP+1 which can be attributed to the fact in this duplex 2AP is not immediately stacked with G, the most efficient quencher amongst all four natural bases. Nevertheless, 2AP in this position is still well-stacked, though it is located at the end of the base-paired portion. In 2AP-9, τ_1 is significantly shorter, this is because 2AP is stacked between two guanines and thus a more quenched environment can be formed, which can also be evidenced by higher A_1 value.

	2AP+1				2AP-1				2AP-9 (control)			
Enzyme	**τ_1/ns (A_1)**	**τ_2/ns (A_2)**	**τ_3/ns (A_3)**	**τ_4/ns (A_4)**	**τ_1/ns (A_1)**	**τ_2/ns (A_2)**	**τ_3/ns (A_3)**	**τ_4/ns (A_4)**	**τ_1/ns (A_1)**	**τ_2/ns (A_2)**	**τ_3/ns (A_3)**	**τ_4/ns (A_4)**
None	0.060 (0.75)	0.30 (0.17)	1.5 (0.06)	5.7 (0.02)	0.036 (0.77)	0.42 (0.10)	2.0 (0.09)	7.1 (0.04)	0.026 (0.85)	0.42 (0.06)	2.0 (0.07)	6.5 (0.02)
WT	0.045 (0.72)	0.55 (0.10)	2.4 (0.16)	5.8 (0.02)	0.053 (0.70)	0.57 (0.15)	2.2 (0.10)	7.1 (0.05)	0.026 (0.83)	0.45 (0.07)	2.0 (0.07)	6.5 (0.03)
Y40A	0.084 (0.57)	0.40 (0.30)	1.6 (0.11)	7.0 (0.02)	0.069 (0.56)	0.72 (0.20)	2.5 (0.16)	7.9 (0.08)				
R100A	0.054 (0.74)	0.38 (0.14)	1.6 (0.10)	7.3 (0.02)	0.070 (0.55)	0.81 (0.20)	2.6 (0.16)	7.9 (0.09)				
K93A	0.060 (0.75)	0.41 (0.12)	1.6 (0.11)	7.0 (0.02)	0.062 (0.60)	0.76 (0.18)	2.5 (0.15)	7.8 (0.07)				

Table 4.2—Lifetime parameters for the 2AP-labelled duplexes and the enzyme-DNA complexes.

Moreover, the graphical representation of decay parameters for free duplex 2AP+1 and 2AP-1, and their respective binary complexes with WT FEN or other three mutants are shown in Figure 4.7 and Figure 4.8.

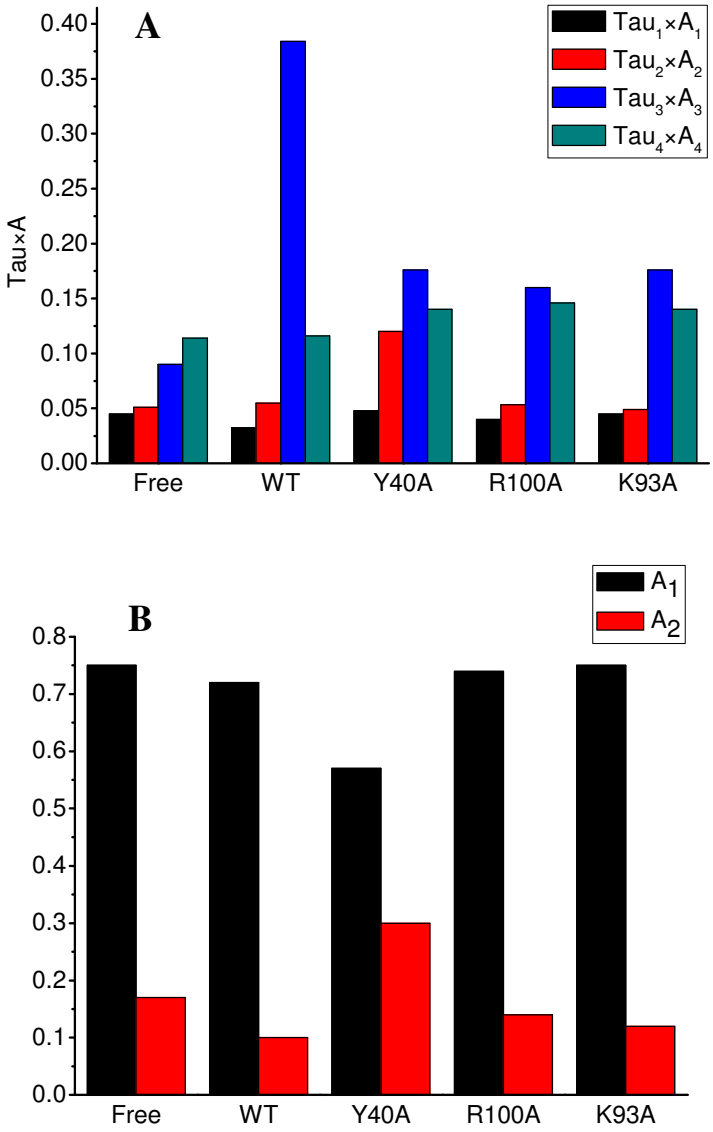


Figure 4.7—Graphical representation of decay parameters for free duplex 2AP+1 and its binary complex with WT hFEN1 or three other mutants. (A) Bar chat representing $\tau_i \times A_i$ values for each of the four components for free 2AP+1 and 2AP+1 in complex with enzymes. (B) Bar chat of A-factors of the first two lifetime components for 2AP+1.

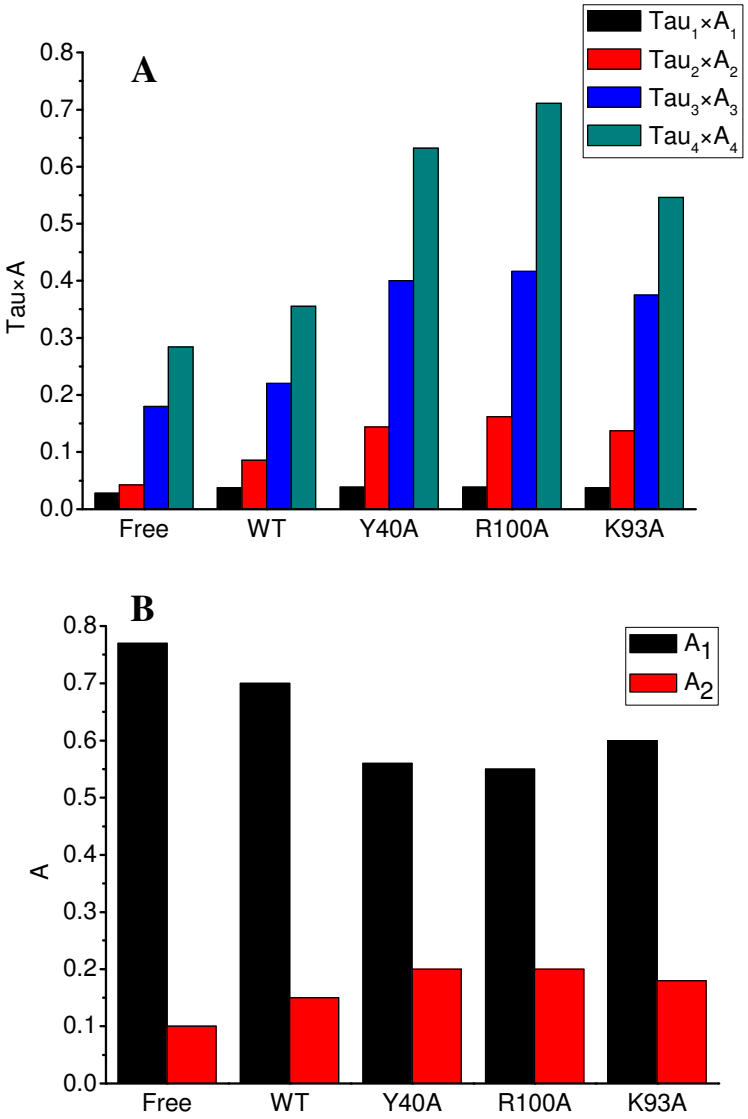


Figure 4.8—Graphical representation of decay parameters for free duplex 2AP-1 and its binary complex with WT hFEN1 or three other mutants. (A) Bar chat representing $\tau_i \times A_i$ values for each of the four components for free 2AP-1 and 2AP-1 in complex with enzymes. (B) Bar chat of A-factors of the first two lifetime components for 2AP-1.

4.3.3 Lifetime parameters of 2AP+1, 2AP-1, 2AP-9 in complex with WT hFEN1

Firstly, as shown in Table 4.2, 2AP-9 shows minimal change in decay parameters on binding of WT enzyme, signifying that no distortion of the duplex occurs at this point distant from the scissile phosphate. In contrast to this, 2AP+1 shows a manifest increase in both the magnitude of 3rd lifetime component, τ_3, and its amplitude A_3. The consequent large increase in the product $A_3 \times \tau_3$, can be clearly seen in Figure 4.7A.

As reported before, empirically the remarkable difference between ssDNA and its corresponding dsDNA lies in the 3rd component and can be well reflected by $A_{3\times}\tau_3$ value. Thus, binding of the WT enzyme causes unzipping in the vicinity of the scissile phosphate and 2AP in the +1 position experiences an environment resembling that in a single strand. However, a highly stacked, quenched conformation, with a high population still remains. Surprisingly the shorter value of τ_1 indicates more rapid quenching than in the unbound duplex. This is contrary to expectation for unzipping of the duplex, but can be attributed to the fact the 2AP lies close to the aromatic ring of nearby Y40 tyrosine residue as shown in Figure 4.9. Previously, it has been found that the fluorescence of 2AP flipped into the active site of methyltransferase TaqI can be efficiently quenched by stacking interactions with tyrosine residue, resulting in sub-nanosecond decay times.[16] Therefore, the lifetimes of 2AP in +1 position is consistent with the fact that Y40 is observed to stack with the +1 adenine base in the crystal structure of hFEN1:substrate DNA.

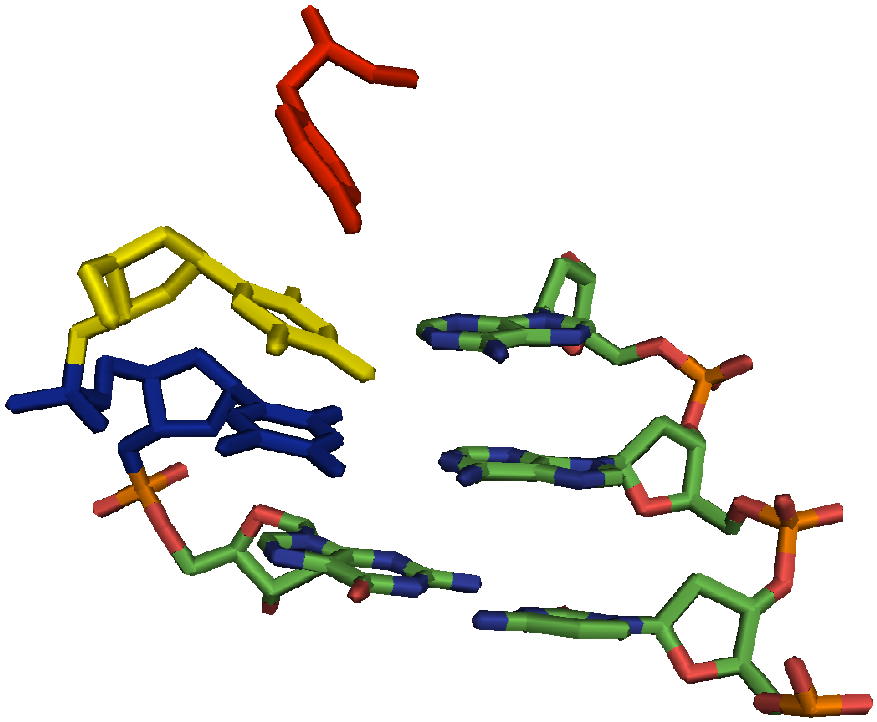


Figure 4.9—Close-up of WT hFEN1's binding to the substrate DNA, which Y40 stacks with nucleobases in +1 and -1 position (PDB: 3Q8L). Tyrosine residues (Y40) is in red, nucleobases at positions -1 and +1 relative to scissile phosphodiester bond are coloured blue and yellow respectively.

As displayed in Figure 4.7A and Figure 4.8A, the response of 2AP-1 to WT hFEN1 binding is different from that of 2AP+1 and indicates that the -1 base is constrained in a well-stacked, quenched, duplex-like environment. The absence of a response that is characteristic of unzipping can be deduced at first glance, but this can also be traced to interaction with Y40, as revealed by comparison with the Y40A complex. It has been observed that there are similar effects of the stacking of 2AP with tyrosine (or tryptophan) residues in enzyme active sites, where the quenching interaction mimics that of interbase interaction in the DNA duplex. Neely *et al.*[12] reported that in some case of interaction of 2AP-labelled DNA and enzyme, the hydrogen bonding interactions between amino acid residues in the enzyme and the 2AP successfully mimic the normal intrahelical environment of the base. The base-flipping event appears to have no significant impact on the conformational structure or dynamics of the duplex.

4.3.4 Lifetime parameters of 2AP+1, 2AP-1 in complex with Y40A hFEN

By comparing with the decay parameters for duplexes in complex with WT hFEN1 and Y40A, it can be clearly seen that removal of the Y40 residue changes the environment of 2AP in both +1 and -1 positions. Comparison with the respective WT complexes shows that in both cases there is a transfer of population from highly quenched to less quenched states, namely the decline in A_1 and increase in A_2, as shown in Figure 4.7B and Figure 4.8B, and an increase in the value of τ_1. Moreover, the environment of 2AP+1 no longer resembles a single strand, reflected by a sharp decrease in $A_3 \times \tau_3$ as shown in Figure 4.7A. All those abovementioned features imply that in complex with Y40A, 2AP in each position of +1 and -1 is rather more weakly stacked and more mobile in conformation than in the free duplex, but there is no evidence for substantial unzipping. Based on the observation that 2AP responds differently in the WT hFEN1 complexes to the Y40A complexes, one can postulate that in the WT complex, both 2AP+1 and 2AP-1 interact with Y40. It is evident from the crystal structure that stacking of the -1 base with Y40 necessitates disruption of the duplex structure as shown in Figure 4.10; this is also consistent with the local unzipping that is reported by 2AP in the +1 position.

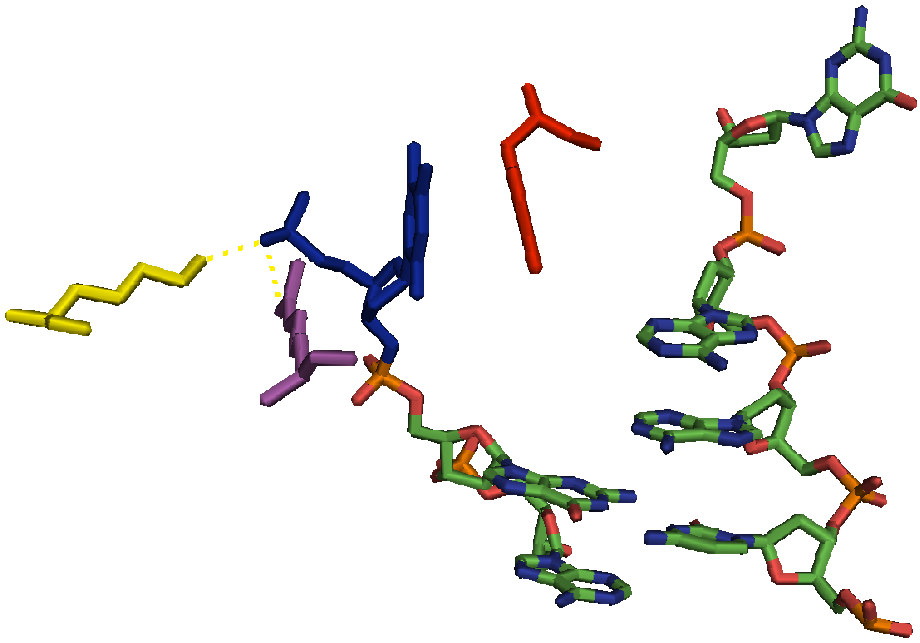


Figure 4.10—Close-up of hFEN1's binding to the -1 phosphate group of the 5' flap in WT hFEN1: product DNA complex (PDB: 3Q8K). The vicinity of three crucial amino acid residues R100 (magentas), K93 (yellow), and Y40 (red) and for catalysis relative to -1 base (blue) are clearly shown. The dash lines between the groups indicate hydrogen binding contact.

4.3.5 Lifetime parameters of 2AP+1, 2AP-1 in complex with R100A/K93A hFEN

The interaction of Y40 with 2AP at the +1 position is apparent in the complexes with R100A and K93A. This can be evidenced by the fact that the A_1, A_2 values are much like those in the WT complex and quite different from those in the Y40A complex, as shown in Figure 4.7B; as well as the similarity of $\tau_1 \times A_1$ and $\tau_2 \times A_2$ values as those of WT complex as shown in Figure 4.7A. However, in contrast to the WT hFEN1 complex, these two mutant (R100A and K93A) complexes do not show significant unzipping of the +1 base, as shown in Figure 4.7A, and in this regard they resemble the Y40A complex in the value of $\tau_3 \times A_3$. The disparity between these complexes and the WT is also evident in the response of 2AP at the -1 position. The decay parameters of 2AP at the -1 position do not resemble those in the WT complex, but are almost identical to those in the Y40A complex for both A-factors and $\tau_i \times A_i$. This indicates there is some enzyme-induced distortion of the duplex, resulting in reduced

interbase stacking, but a duplex-like structure is retained. Also, it signifies that 2AP in -1 position does not stack with Y40 in these two mutant enzymes (R100A and K93A), supporting the lack of unzipping shown by the 2AP+1 response. This implies that the presence of both R100 and K93 residues is necessary for the unzipping of the duplex structure by the WT hFEN1 for nucleotides in both +1 and -1 positions. This is consistent with the biochemical data reported by Tsutakawa *et al.*[7] Lys93 (K93) and Arg100 (R100) interact with the product 5'-phosphate monoester *via* hydrogen binding as shown in the crystal structure (Figure 4.10) and probably stabilise the unzipped conformation of DNA substrates, acting as electrostatic catalysts during hydrolysis. In the K93A and R100A mutants, the rate of the hFEN1-catalysed reaction is drastically reduced by >400-fold.

4.3.6 Summary

Based the time-resolved fluorescence results along with the available crystal structures, the following whole picture emerges, in terms of the conformational dynamics of DNA when binding to the enzyme in solution. In the WT complexes, the +1 and -1 bases are unzipped. The -1 base is displaced from its well-stacked position in the duplex and lies closely with the Y40 residue, also burying in the enzyme, which results in an apolar environment for -1 base resembling it in the B-form DNA. The +1 base is relatively unconstrained in the single strand on the 5' side of the scissile phosphate, but is also subject to the stacking interaction with Y40. The crystal structure of the product complex shows the -1 base stacked with Y40 after cleavage, but we now have evidence that this base stacks with Y40 in order to enter the active site and bind to the metal ions prior to cleavage. The interaction of Y40 with the -1 base stabilises the unzipped conformation; when Y40 is removed from the WT enzyme, the -1 and +1 bases revert to a weakly stacked, duplex-like conformation. The R100 and K93 residues are also involved in unzipping of the duplex, probably by stabilising the unzipped conformation of DNA substrates. In the absence of either of these residues, the -1 base is not stacked with Y40, but AP+1 do see the quenching influence of Y40. This is consistent with the interaction of the +1 base with Y40 on its 5' side as they are in close vicinity, in the absence of unzipping

of the duplex, as seen in the crystal structure of WT hFEN1 in complex with its substrate.

4.4 References

1. T. A. Ceska, J. R. Sayers, G. Stier and D. Suck, *Nature*, 1996, **382**, 90-93.
2. Y. Liu, H.-I. Kao and R. A. Bambara, *Annu. Rev. Biochem.*, 2004, **73**, 589-615.
3. G. K. Schroeder, C. Lad, P. Wyman, N. H. Williams and R. Wolfenden, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 4052-4055.
4. C. G. Tomlinson, K. Syson, B. Sengerová, J. M. Atack, J. R. Sayers, L. Swanson, J. A. Tainer, N. H. Williams and J. A. Grasby, *J. Biol. Chem.*, 2011, **286**, 30878-30887.
5. L. Zheng, J. Jia, L. D. Finger, Z. Guo, C. Zer and B. Shen, *Nucleic Acids Res.*, 2011, **39**, 781-794.
6. L. N. Tumey, D. Bom, B. Huck, E. Gleason, J. Wang, D. Silver, K. Brunden, S. Boozer, S. Rundlett, B. Sherf, S. Murphy, T. Dent, C. Leventhal, A. Bailey, J. Harrington and Y. L. Bennani, *Bioorg. Med. Chem. Lett.*, 2005, **15**, 277-281.
7. S. E. Tsutakawa, S. Classen, B. R. Chapados, A. S. Arvai, L. D. Finger, G. Guenther, C. G. Tomlinson, P. Thompson, A. H. Sarker, B. Shen, P. K. Cooper, J. A. Grasby and J. A. Tainer, *Cell* 2011, **145**, 198-211.
8. L. D. Finger, M. S. Blanchard, C. A. Theimer, B. Sengerová, P. Singh, V. Chavez, F. Liu, J. A. Grasby and B. Shen, *J. Biol. Chem.*, 2009, **284**, 22184-22194.
9. J. A. Grasby, L. D. Finger, S. E. Tsutakawa, J. M. Atack and J. A. Tainer, *Trends Biochem. Sci.*, 2012, **37**, 74-84.
10. E. Y. M. Bonnist and A. C. Jones, *ChemPhysChem*, 2008, **9**, 1121-1129.
11. E. Y. M. Bonnist, K. Liebert, D. T. F. Dryden, A. Jeltsch and A. C. Jones, *Biophys. Chem.*, 2012, **160**, 28-34.
12. R. K. Neely, D. Daujotyte, S. Grazulis, S. W. Magennis, D. T. F. Dryden, S. Klimasauskas and A. C. Jones, *Nucleic Acids Res.*, 2005, **33**, 6953-6960.
13. R. K. Neely and A. C. Jones, *J. Am. Chem. Soc.*, 2006, **128**, 15952-15953.

14. R. K. Neely, S. W. Magennis, S. Parsons and A. C. Jones, *ChemPhysChem*, 2007, **8**, 1095-1102.
15. R. K. Neely, G. Tamulaitis, K. Chen, M. Kubala, V. Siksnys and A. C. Jones, *Nucleic Acids Res.*, 2009, **37**, 6859-6870.
16. T. Lenz, E. Y. M. Bonnist, G. Pljevaljčić, R. K. Neely, D. T. F. Dryden, A. J. Scheidig, A. C. Jones and E. Weinhold, *J. Am. Chem. Soc.*, 2007, **129**, 6240-6248.

Chapter 5---Investigation of EcoP15I Base Flipping Mechanism and its Matched and Mismatched DNA Cleavage

5.1 Introduction

5.1.1 Introduction

The enzyme used in this investigation was EcoP15I, a type III R-M system.[1,2] EcoP15I is a multimetric enzyme composed of two types of subunit: restriction (Res) subunit and modification (Mod) subunit. As an SAM, (AdoMet)-dependent DNA methyltransferase, EcoP15I specifically binds to the 5'-CAGCAG-3' recognition sites and transfers a methyl group to the second adenine in the presence of magnesium and SAM.[3] When performing restriction activity, EcoP15I requires two unmethylated inversely oriented asymmetric recognition sequences. The enzyme displays a preference for sites in a head-to-head arrangement (as shown in Figure 5.1). The cleavage reaction requires the hydrolysis of ATP. The 'head' of the sequence can be defined as the dG on the 3' end of the CAGCAG on the top strand. The DNA cleavage efficiency is affected by the distance of the two recognition sites. DNA cleavage takes place on the 3' side of either recognition site 25-27 base pairs downstream of the recognition site, irrespective of the length of the DNA between the sites, to generate a single-strand protrusion of two nucleotides.[4,5]

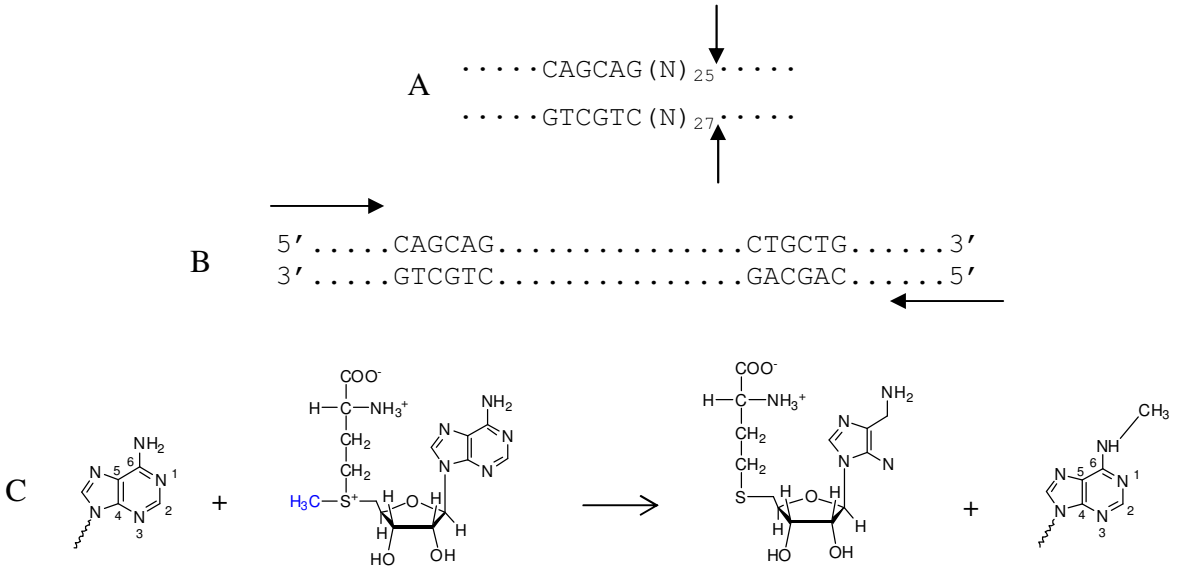


Figure 5.1—Recognition sequence (A), head to head recognition site arrangement (B) and methyl group transfer reaction of EcoP15I (C). The vertical arrows indicate the cleavage sites.

Wagenführ *et al.*[6] used limited proteolysis, mass spectrometry and insertional mutagenesis to characterise the structural domains in the type III restriction endonuclease EcoP15I. They identified at least two domains in the restriction subunits; an N-terminal domain (amino acids 40-705) of the Res subunit motor domain or translocase domain (Tr) containing conserved helicase motifs and a C-terminal endonuclease domain (amino acids 729-966) harbouring the active site responsible for DNA cleavage. Both the above mentioned structural domains are connected *via* a flexible linker region spanning 23 amino acid residues. These two domains are highlighted in Figure 5.1.

It has been reported that EcoP15I DNA methyltransferase introduces a large structural distortion within the recognition sequence upon binding. Reddy *et al.*[7] measured the steady-state fluorescence intensity of 2AP substituted duplexes in the absence or presence of enzyme and cofactor. An eight to ten-fold fluorescence enhancement was observed upon EcoP15I binding to its cognate duplexes in the presence of cofactor SAM cofactors analogues such as sinefungin (Sf), SAH.

Moreover, the fluorescence change only happens for the bases in the recognition sites rather than the bases outside. The fluorescence intensity observed upon enzyme binding was much more significant than the fluorescence of the single-stranded 2AP oligonucleotide. This observation indicates that the fluorescent change was not merely an enzyme-induced local base disturbance of the double helical structure resulting in a single-stranded oligonucleotide region with 2AP present. Reddy *et al.*[7] concluded that the two adenines in the recognition sites rotate to become structurally unstacked. Indeed, this is a clear indication that that EcoP15I MTase uses the base flipping mechanism for achieving adenine methyl group transfer.

Thus, an increase in steady-state intensity is a useful indicator for determining the destacking movement of 2AP in the duplex. However, there is a lack of information concerning the conformation distribution of 2AP, which makes it difficult to analyse the intensity change for a DNA-enzyme complex. Different distributions of the 2AP population amongst the various conformational states of the duplex-enzyme complex could give similar steady-state fluorescence intensities.

>R.EcoP15I 970 aa Restriction subunit M.W.=110956.4922

1-50 MSKGFTLEKN LPHQKAGVDA VMNVFVSATP HLTDNVAVRL LANPELKLSE
51-100 QQYYNNIKNV QAFNGIAHSK DNHNAKSNII DVSMETGTGK TYTYIKTIFD
101-150 LNKSFGINKF IIIVPTLSIK AGTVNFLKSD ALKEHFRDDY KRELRTYVVE
151-200 SQKNAGKNTK SYMPQAIHDF VEASNFNKKY IHVLVINSGM INSKSLTDTY
201-250 DTGLLDNQFN TPVDALRAVK PFIIIDEPHR FPTGKKTWEN IEKFNAQYII
251-300 RYGATFSEGY KNLVYRLTAV DAFNDDLVKG IDAYIEDIVG DGNANLKFVK
301-350 SDGKEATFEL NENNNKKSFK LAKGESLSKT HSAIHDLTLD ALNKSTAVLS
351-400 NGIELKIGSS INPYSYDQTL ADNMMRKAVK EHFKLEKELL TQRPRIKPLT
401-450 LFFIDDIEGY RDGNDISGSL KTKFEEYVLA EANELLKTEQ DAFYKNYLEK
451-500 TVTNISSVHG GYFSKDNSDK DDKIEQEINE ILHDKELLLS LDNPRRFIFS
501-550 KWTLREGWDN PNVFQICKLR SSGSTTSKLQ EVGRGLRLPV NEYMCRVKDR
551-600 NFTLKYYVDF TEKDFVDSLV KEVNESSFKE RVPSKFTQEL KEQIMAQYPE
601-650 LSSRALMNEL FNDEIIDDND NFKDSDAYSR LKSKYPAAFP IGVKPGKIKK
651-700 ATDGKRRTKM RVGKFSELKE LWDLINQKAV IEYKINSESE FLSIFKSFML
701-750 EETERFTKSG VHTRIDKIYI HNDMAMSKSI VSDDDDFAKL NTMSYREFLD
751-800 NLSQTIFVKH GTLHKVFCDI KDTINITEYL NIQTIRKIKS GFSKYLLNNS
801-850 FNKFSLGYNL ISGSIHPTKF TNADGNPLGE VLSSDLGVLQ DNAKAPLDTY
851-900 LFEEVFYDSE LERRNITDRE IQSVVVFSKI PKNSIKIPVA GGYTYSPDFA
901-950 YVVKTAEGDY LNFIIETKNV DSKDSLRLEE KRKIEHAQAL FNQISQSVKV
951-970 EFRTQFANDD IYQLIKSALP

>M.EcoP15I 645 aa Modification subunit M.W.=74221.9219

MKKETIFSEV ETANSKQLAV LKANFPQCFD KNGAFIQEKL LEIIRASEVE
LSKESYSLNW LGKSYARLLA NLPPKTLLAE DKTHNQQEEN KNSQHLLIKG
DNLEVLKHMV NAYAEKVKMI YIDPPYNTGK DGFVYNDDRK FTPEQLSELA
GIDLDQAKRI LEFTTKGSSS HSAWLTFIYP RLYIARELMR EDGTIFISID
HNEFSQLKLV CDEIFGEQNH VGDLVWKNAT DNNPSNIAVE HEYIIVYTKN
KEQLISEWKS NISDVKNLLV NIGEEFASKY TGNELQEKYT QWFREHRSEL
WPLDRYKYID KDGIYTGSQS VHNPGKEGYR YDIIHPKTKK PCKQQPLMGY
RFPLDTMDRL LSEEKIIFGD DENKIIELKV YAKDYKQKLS SVIHLDGRVA
TNELKELFPE MTQPFTNAKT IKLVEDLISF ACDGEGIVLD FFAGSGTTAH
TVFNLNNKNK TSYQFITVQL DEPKKDKSDA MKHGYNTIFD LTKERLIRAS
KKNRDQGFKV YQLMADFRAK DESELTLSNH TFFDDVVLTP EQYDTLLTTW
CLYDGSLLTT PIEDVDLGGY KAHLCDGRLY LIAPNFTSEA LKALLQKVDS
DKDFAPNKVV FYGSNFESAK QMELNEALKS YANKKSIELD LVVRN

Table 5.1—The amino acid sequences and calculated molecular mass of Res and Mod subunits of EcoP15I. In Res subunits, the translocase domain (Tr) is highlighted in blue and the endonuclease domain is highlighted in red.

5.1.2 Aims of this chapter

In this chapter, the author intended to verify the base flipping mechanism hypothesis adopted by EcoP15I methytransferase using time-resolved fluorescence. Initially, steady-state measurements were also conducted with the purpose of testing if Reddy *et al.*'s results could be repeated. Additionally the steady-state data can help to interpret lifetime data and the comparison of both would be more informative and conclusive. So far no crystal structure for EcoP15I or its binary complex with DNA (EcoP15I-DNA) is available. Thus, a fluorescence-based approach would be a very practical way to investigate the interaction of EcoP15I and substrate DNA. The author expected to discover more information concerning structural changes to the duplex upon enzyme binding.

As reported by Buchner *et al.*[8], endonuclease EcoP15I can be used as tool for counting CAG repeats. An expansion in the number of CAG repeats located in the first exon of the Huntingtin gene (*HTT*) is known to be responsible for HD, which is a progressive neurodegenerative disorder with autosomal-dominant inheritance.[9,10] It has been shown that DNA fragments containing super-long Huntingtin CAG repeats adopt unusual structures, which give rises to A·A and T·T mismatches.[11,12] It is therefore interesting to examine whether a duplex containing these mismatches can act as a substrate for EcoP15I. Or, fundamentally, what is the cutting pattern of endonuclease EcoP15I on a duplex DNA with regard to the arrangement of two adjacent sites on its substrate DNA duplex? The author examined this point using nine 120-mer duplexes.

5.2 Experimental

5.2.1 Type III restriction endonuclease EcoP15I

EcoP15I (2500 units/mL, New England Biolabs, MA, USA) was used as supplied. EcoP15I consists of methylation (Mod) and restriction (Res) subunits. Each Res subunit (GenBank ID: CAG24073.1) is composed of 970 amino acid residues and each Mod subunit (GenBank ID: CAA29616.1) is composed of 645 amino acid residues. The molecular weight for Res and Mod subunits is 110957 Da and 74222 Da respectively ExPASy-ProtParam tool[13,14] at http://web.expasy.org/cgi-bin/protparam/protparam was used to calculate the extinction coefficient values. The extinction coefficient of EcoP15I at 280 nm can be calculated as 81600 M^{-1} cm^{-1} for Res and 81710 M^{-1} cm^{-1} for Mod, assuming all cysteine residues were reduced.

5.2.2 Oligonucleotides used in this chapter

EcoP15I-2AP1-TT and EcoP15I-2AP2-TT are the duplexes with 2AP base substituting the second and first adenine base within the EcoP15I recognition site (5'-CAGCAG-3') respectively. EcoP15I-2AP3-TT was used as a control sequence in which the 2-aminopurine base is placed outside the recognition site. All three duplexes were used in the EcoP15I-DNA interaction study, in order to investigate whether EcoP15I is able to flip the adenine base(s), using both steady-state and time-resolved fluorescence measurements. When annealing the 2AP-labelled strand with its unlabelled complementary one, the ratio of the latter to the former used was 1.2:1 to assure the former was completely double-stranded. All the duplexes were annealed by heating two complementary strands at 90°C for 5 mins and cooled down overnight. Table 5.3 lists ten 120-mer long single-strand oligonucleotides which were designed based on the paper published by Mücke *et al.*[4] with modification. These oligonuclesotides were all used to form duplexes whose sequences are shown in Table 5.4. The extinction coefficient at 260 nm was used for calculating the exact concentration of the DNA. The molar ratio of two complementary strands for

annealing was 1:1, which was done by heating the sample up at 90°C for 5 mins and cooled down overnight. Extinction coefficient was calculated using Integrated DNA Technologies biophysics online software[15] (http://biophysics.idtdna.com/).

EcoP15I-2AP1-TT	(5′-3′)TAGGTCAGAATT**CAGC2G**ACCCTAAGTAGCC (3′-5′)ATCCAGTCTTAA**GTCGTC**TGGGATTCATCGG
EcoP15I-2AP2-TT	(5′-3′)TAGGTCAGAATT**C2GCAG**ACCCTAAGTAGCC (3′-5′)ATCCAGTCTTAA**GTCGTC**TGGGATTCATCGG
EcoP15I-2AP3-TT	(5′-3′)TAGGTC2GAATT**CAGCAG**ACCCTAAGTAGCC (3′-5′)ATCCAGTCTTAA**GTCGTC**TGGGATTCATCGG

Table 5.2—Names and sequences of DNA duplexes used in the steady-state and time-resolved lifetime measurements. The underline region of EACH duplex is the recognition site of EcoP15I and 2 stands for 2AP DNA analogue base.

Name	Sequence (5'-3')	Extinction coefficient at 260 nm Liter/(mol·cm)
B1283	GGTACCCGGGGATCAGCAGATCCAGGTCAGCCGTGATCCTCCATGACGAGTCGCTAAATT GTTATGGAATCTGTTCATGGAGAGACGAGGACGACCTGGAGCTGCTGATTGGCGTAATCA	1167500
B1284	GGTACCCGGGGATCAGCAGATCCAGGTCAGCCGTGATCCTCCATGACGAGTCGCTAAATT GTTATGGAATCTGTTCATGGAGAGACGAGGACGACCTGGAGCAGCAGATTGGCGTAATCA	1178100
B1285	TGATTACGCCAATCAGCAGCTCCAGGTCGTCCTCGTCTCTCCATGAACAGATTCCATAAC AATTTAGCGACTCGTCATGGAGGATCACGGCTGACCTGGATCTGCTGATCCCCGGGTACC	1124700
B1286	TGATTACGCCAATCAGCAGCTCCAGGTCGTCCTCGTCTCTCCATGAACAGATTCCATAAC AATTTAGCGACTCGTCATGGAGGATCACGGCTGACCTGGATCAGCAGATCCCCGGGTACC	1135300
B1503	TGATTACGCCAATCTGCTGCTCCAGGTCGTCCTCGTCTCTCCATGAACAGATTCCATAAC AATTTAGCGACTCGTCATGGAGGATCACGGCTGACCTGGATCAGCAGATCCCCGGGTACC	1124700
B1696	TGATTACGCCAATCAGCTGCTCCAGGTCGTCCTCGTCTCTCCATGAACAGATTCCATAAC AATTTAGCGACTCGTCATGGAGGATCACGGCTGACCTGGATCAGCAGATCCCCGGGTACC	1130000
B1697	TGATTACGCCAATCTGCAGCTCCAGGTCGTCCTCGTCTCTCCATGAACAGATTCCATAAC AATTTAGCGACTCGTCATGGAGGATCACGGCTGACCTGGATCAGCAGATCCCCGGGTACC	1130000
B1791	GGTACCCGGGGATCAGCAGATCCAGGTCAGCCGTGATCCTCCATGACGAGTCGCTAAATT GTTATGGAATCTGTTCATGGAGAGACGAGGACGACCTGGAGCAGCTGATTGGCGTAATCA	1172800
B1792	GGTACCCGGGGATCAGCTGATCCAGGTCAGCCGTGATCCTCCATGACGAGTCGCTAAATT GTTATGGAATCTGTTCATGGAGAGACGAGGACGACCTGGAGCAGCTGATTGGCGTAATCA	1167500
B1793	TGATTACGCCAATCAGCAGCTCCAGGTCGTCCTCGTCTCTCCATGAACAGATTCCATAAC AATTTAGCGACTCGTCATGGAGGATCACGGCTGACCTGGATCAGCTGATCCCCGGGTACC	1130000

Table 5.3—Single-stranded oligonucleotides used for making DNA duplexes in EcoP15I matched and mismatched duplex cutting experiment.

Table 5.4—Names and sequences of dsDNA substrates for EcoP15I duplex cleavage experiments. The based underlined are the bases of interest. The red and green boxes highlight the sequences within the recognition sites. The vertical arrows indicate the cutting positions for each duplex and the pink arrows indicate the mismatch in the region concerned.

5.2.3 Size-exclusion chromatography analysis of EcoP15I

An analytical HPLC gel filtration column calibrated with protein standards (apoferritin 443 kDa, β-amylase 200 kDa, alcohol dehydrogenase 150 kDa, bovine serum albumen 66 kDa and carbonic anhydrase 29 kDa) was used to determine the molecular weight of EcoP15I in a buffer composed of 20 mM Tris-HCl, 20 mM 2-(N-morpholino)ethanesulfonic acid (MES), 10 mM magnesium chloride, 7 mM β-mercaptoethanol, 200 mM sodium chloride, 0.1 mM EDTA, pH 6.5 at room temperature as previously described[16]. The flow rate of mobile phase was 0.5 mL/min and the elute was monitored at wavelength 350 nm continuously using a flow-through fluorescence detector excited at 295 nm. The EcoP15I was tested at concentrations from 3840 nM to 2.56 nM.

5.2.4 SDS-PAGE analysis of EcoP15I

SDS-PAGE of EcoP15I was done as described in Chapter 2. 0.375, 0.75, 0.4, 0.5 μL of EcoP15I samples (34.3 μM as Res_1Mod_2 stoichiometry) were loaded into lane 1 to lane 4 respectively, protein ladder was loaded in the rightmost lane.

5.2.5 Densitometry of SDS-PAGE for determining the stoichiometry of EcoP15I

NIH ImageJ software was used as previously described.[17] NIH ImageJ software was firstly calibrated with Kodak No. 3 calibrated step tablet with a density ranging from 0.05 to 3.05 OD. The absolute density value of these two bands was generated by this software and subsequently the molar concentration of the two subunits can be then calculated based on the following equation:

$$\frac{\text{Band density value}}{\text{molecular weight of individual subunit}}$$

5.2.6 Steady-state fluorescence measurements

Three synthesised 2AP-labelled oligonucleotides (shown in Table 5.2) were annealed with their respective complementary strands and used for investigating the interaction of EcoP15I with DNA. Each duplex was buffered with 50 mM Tris-HCl, 20 mM sodium chloride, 10 mM magnesium chloride, pH 7.5. The cofactors, SAH and SAM were both in 100 μM concentration where needed. 200 μL samples of 500 nM duplex plus 2.5 μM EcoP15I (Res_1Mod_2 stoichiometry) were incubated for 30 mins at 25°C prior to measurement. Steady-state fluorescence spectra were measured using a FluoroMax (Horiba Jobin Yvon) photon counting spectrofluorometer. Spectra were recorded with a bandpass of 10 nm for both excitation (315 nm) and emission (370 nm) as previously described with modification.[18]

5.2.7 Time-resolved fluorescence measurements

Time-resolved fluorescence data were measured as described in Chapter 2. Fluorescence decay curves were collected at an excitation wavelength 315 nm and three diverse emission wavelengths (365, 380, 395 nm) as previously described with modification[19]. When measuring the duplexes in the absence of EcoP15I, the concentration of each duplex was 10 μM. When measuring EcoP15I-2AP1-TT+EcoP15I, the concentration for EcoP15I-2AP1-TT was 0.5 μM, the concentration of EcoP15I was 2.5 μM in order to make sure DNA was fully bound. When measuring EcoP15I-2AP2+EcoP15I and EcoP15I-2AP2+EcoP15I, the duplexes and EcoP15I used were 10 μM respectively.

5.2.8 Polyacrylamide gel for visualising DNA cleavage by EcoP15I

Unlabelled duplexes as shown in Table 5.4 were used in this experiment at a concentration of 4 μM. 5 μL of EcoP15I stock solution (34.3 μM) was added to give a final EcoP15I concentration of 3.12 μM. All samples were incubated at 37°C for ~20 hours in the reaction buffer supplemented with 5 μM ATP and 0.2 mg/mL bovine serum albumin (BSA). The reaction was then run on a 15% polyacrylamide gel in

1×Tris/Borate/EDTA buffer at 70 V for 4 hours and cooled with ice in the course of gel running to keep the temperature low. The gel was subsequently stained with 0.5 μg/mL ethidium bromide for 0.5 hours and viewed under UV light. The reaction buffer for EcoP15I DNA cleavage was 100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM dithiothreitol (DTT), pH 7.9, at room temperature.

5.3 Results and Discussion

5.3.1 Size-exclusion chromatography and quantitative densitometry of EcoP15I

EcoP15I enzyme samples of different concentration ranging from 3840 nM to 2.56 nM were tested and the elution profile for EcoP15I at a concentration of 1280 nM is shown in Figure 5.2. A sharp symmetric single peak having a retention time of roughly 5.76 minutes was observed, corresponding to a molecular mass of ~ 310 kDa. The apparent molecular mass *vs.* concentration is shown in Figure 5.3. It can be seen that the molecular mass is relatively constant over the concentration range used in this experiment. The average molecular mass indicated by gel filtration was 307±6.6 kDa. Res_1Mod_2 (calculated M.W.=260 KDa), Res_2Mod_1 (calculated M.W.=296 kDa) Res_2Mod_2 (calculated M.W.=370 KDa) all give an approximately molecular mass resembles this value. There exists a discrepancy between molecular mass of EcoP15I derived from gel filtration HPLC experiment and the calculation from the amino acid composition. However, the values obtained from gel filtration are only accurate for completely globular protein complexes. Hence, a discrepancy is to be expected. Moreover, it can also be inferred from the experimental data that the EcoP15I may adopt an elongated structure in solution.

Additional quantitative densitometry experiment was done to settle this point. Firstly, as shown in Figure 5.4, the SDS-PAGE clearly demonstrates the high purity of EcoP15I. The gel result shows two sharp bands after Coomassie Blue staining at molecular weight of ~110 kDa and ~75K kDa respectively, close to those expected from the amino acid sequence of two subunits of EcoP15I.

The molarity ratio of the two subunits was quantified *via* densitometry assuming that a linear relationship between protein mass and colour intensity. The two well-resolved peaks in each densitometry picture (Figure 5.5) correspond two subunits in each line of the SDS-PAGE (Figure 5.4). The molar ratio of Mod and Res subunits was calculated accordingly: lane 1 (1:2.35); lane 2 (1: 1.84); lane 3 (1:2.36); lane 4 (1:2.07). The ratio was 2.155±0.250 on average. The author performed the densitometry measurements at various concentrations and all results showed minimal deviation. Based on these results a Res_1Mod_2 stoichiometry can be confirmed.

The stoichiometry of EcoP15I had been reported earlier as a stable Mod_2Res_2 hetero-tetramer.[5] It has also been shown that M.EcoP15I occurs as a dimer over a wide range of concentrations by both gel filtration and chemical cross-linking experiments.[20] In this chapter, the subunit stoichiometry has to be confirmed in order to accurately calculate enzyme concentration.

Although our results are in conflict with some previous publications, they nevertheless agree perfectly with a very recent data shown by Wyszomirski *et al.*[21]. In their paper, they used analytical ultracentrifugation, densitometric analysis of SDS-PAGE, second-derivative UV absorption spectroscopy to confirm a Res_1Mod_2 stoichiometry for EcoP15I. It was thought by Wyszomirski *et al.*[21] the reason for previous publication claiming the Res_2Mod_2 tetramer was mainly due to higher aggregates and excess subunits. These explanations can be well-proved by the experiments described in Wyszomirski *et al.*'s paper, where they added an additional size-exclusion step in protein purification and consequently, two distinct peaks were obtained. The first with short retention time mainly contained the Res_2Mod_2 fraction; whereas the longer retention time peak mostly contained the Res_1Mod_2 fraction. Determining the restriction activity enabled them to find that the Res_1Mod_2 has a ~6-fold higher activity than Res_2Mod_2 and therefore Res_1Mod_2 is the active EcoP15I REase *in vitro*.[21]

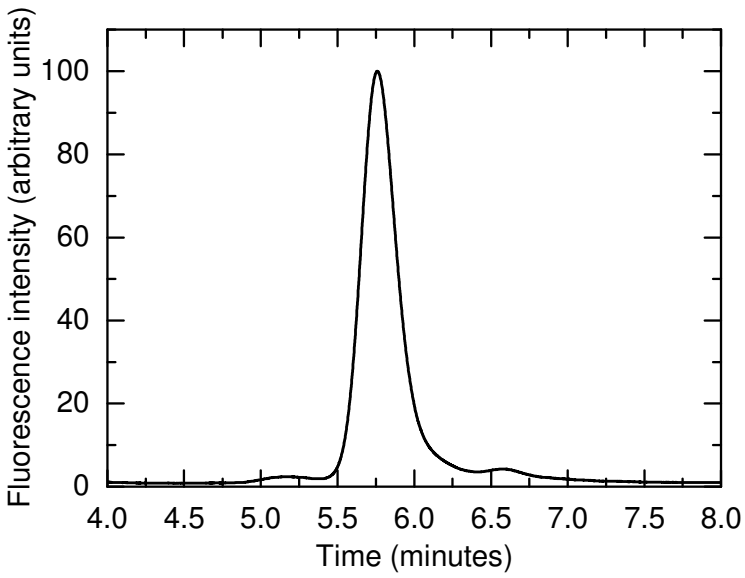


Figure 5.2—Size-exclusion chromatography to investigate the molecular mass of EcoP15I in solution. The elution profile (1280 nM shown) shows a main single (and two negligible shoulder peaks) peak at ~5.76 minutes corresponding to a molecular mass of 310 kDa.

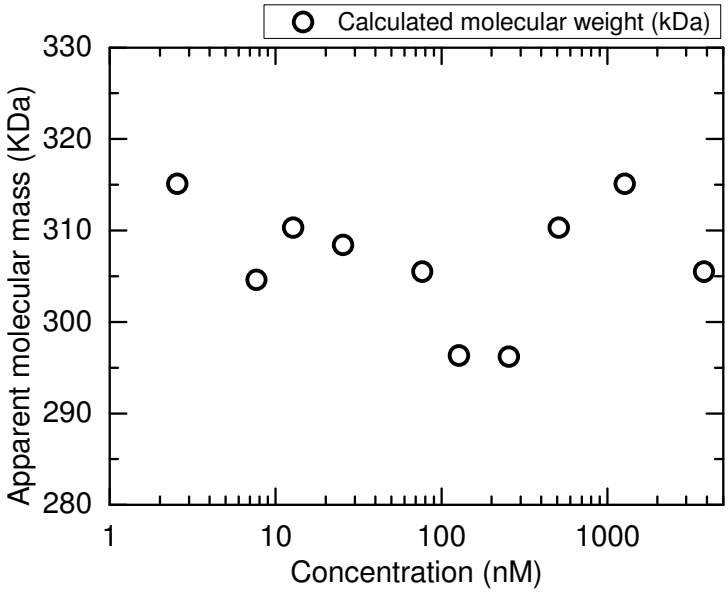


Figure 5.3—Dependence of molecular mass deduced from the single main peak in the size-exclusion HPLC experiments (as exemplified in Figure 5.2) as a function of protein concentration injected onto the column.

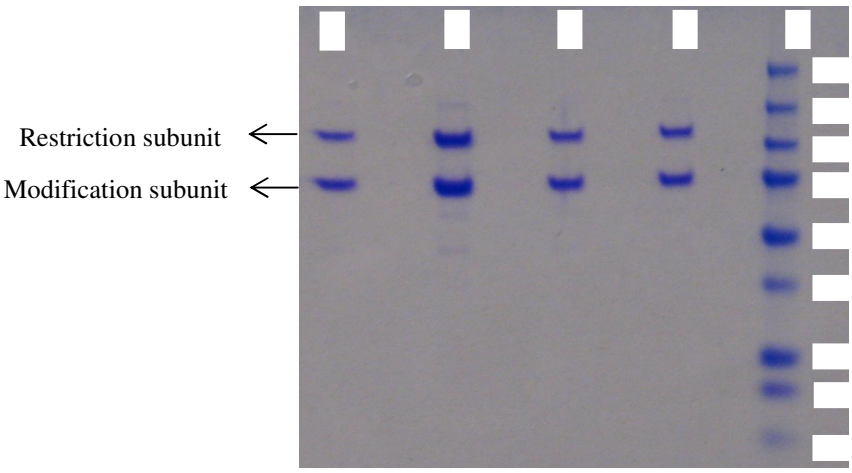


Figure 5.4—SDS-PAGE analysis of EcoP15I enzyme. From the gel, it can be seen a high purity of EcoP15I. The arrow indicates the restriction subunit and modification subunits for this heterogeneous enzyme.

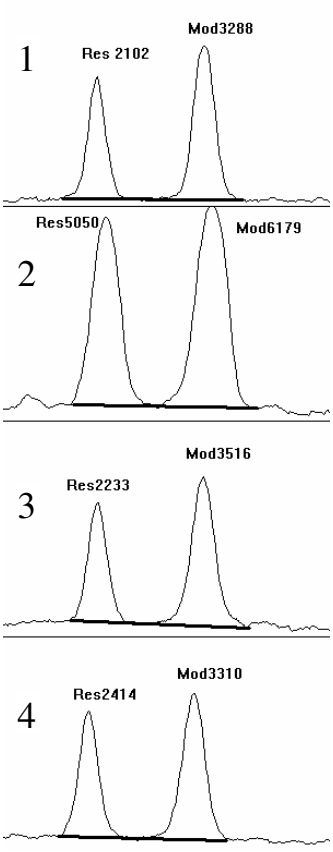


Figure 5.5—Quantitative densitometry analysis of EcoP15I SDS-PAGE shown in Figure 5.4. The peaks represent the density of Res and Mod subunits band in the SDS-PAGE. 1, 2, 3, 4 are the densitometry results derived from the lane 1, 2, 3, 4 in SDS-PAGE shown in Figure 5.4. The molar ratios of Res subunit to Mod subunit for each lane are listed on the right.

5.3.2 Steady-state fluorescence measurements

The excitation wavelength used was 315 nm for all the measurements. The emission spectra of buffer employed are shown in Figure 5.6 f, in which the Raman scattering can be identified as the hump peaking at ~370 nm. The other spectra are all shown after subtraction of the buffer background. The emission spectrum of each ssDNA (Figure 5.6e, 5.7d and 5.9c) show a typical peak at ~380 nm. The fluorescence intensity of 2AP-labelled ssDNA was 3 to 6-fold higher than the corresponding dsDNA. Because 2AP in a ssDNA is less stacked, it gives a higher quantum yield in solution. For the emission of the 2AP-labelled duplex, all exhibit dual humps, which are anomalous. The hump at ~380 nm is normal, while the hump at longer wavelength (445 nm) is unusual. This may be due to sequence context of 2AP which is in a guanine-rich region and thus having a more stacked conformation. The perfect π-π stacking can allow a long wavelength emission as reported before.[22-24]

As shown in Figure 5.6, for the EcoP15I-DNA binary complex, EcoP15I-DNA-SAM, and EcoP15I-DNA-SAH ternary complexes, the fluorescence intensity increases significantly compared with free duplex. Amongst the three complexes, as shown in Figure 5.6 a-5.6 c, the binary complex, whose intensity is as 23.7-fold as free duplex, is the lowest enhancement compared with free duplex. EcoP15I-DNA-SAH has 32.2-fold as free duplex in intensity, which gives the highest enhancement in comparison with free duplex. The EcoP15I-DNA-SAM shows a 24.6-fold increase in fluorescence intensity as free duplex.

Since 2AP in EcoP15I-2AP3-TT duplex is located outside the recognition site, the enzyme is not able to flip this base. Thus the slight increase in the fluorescence must be due to a minor disturbance of 2AP upon binding for binary complex for EcoP15I-2AP3-TT as shown in Figure 5.9. The same fluorescence increase pattern can be observed for duplex EcoP15I-2AP2-TT and also for its binary and ternary complexes as shown in Figure 5.7. From the close-up spectra as shown in Figure 5.8 and 5.10, the binary and ternary spectra of EcoP15I-2AP2-TT and binary complex of EcoP15I-2AP3-TT lose the long wavelength emission peak. The fluorescence

intensity increases slightly at 380 nm emission for both EcoP15I-2AP2-TT and EcoP15I-2AP3-TT duplexes upon enzyme binding, which is due to the fact that the binding of enzyme results in destacking of 2AP base.

The absence of DNA-enzyme co-crystal structures encourages the development of solution-based approaches to determine the base flipping interaction. One strategy involves replacing the target base, usually the adenine base with a fluorescent analogue, 2AP. When highly stacked in a dsDNA, 2AP is strongly quenched through base stacking thus showing very low fluorescence intensity. Destacking of 2AP loses this highly stacked conformation and results in an increase in fluorescence intensity. For base flipping, it involves a rotation of phosphate backbone bond in dsDNA to expose an out-of-stack base, which can be used subsequently as a substrate for an enzyme-catalyzed chemical reaction.[25] If the target base is replaced by 2AP, the enhancement in intensity can be taken as an indicator of base flipping. This method has been applied to a number of MTases, most of which induce about a 10-fold or more increase in intensity of the 2AP target on enzyme binding.[26-30]

If the increased 2AP fluorescence intensity observed upon enzyme binding is much more substantial than the fluorescence of the 2AP-labelled ssDNA, this indicates a much less stacked conformation of 2AP than in single-stranded form. Thus this kind of de-stacking is not just an enzyme-induced local unwinding of the helical structure resulting in a region of ssDNA surrounding the 2AP. Therefore the significant enhancement for EcoP15I-2AP1-TT carrying the 2AP base in the target methyl transfer site (second adenine) of the recognition sequence upon enzyme binding, together with the fact that this kind of increase is much more intense than the fluorescence of the 2AP-labelled ssDNA, firmly indicates that EcoP15I adopts a base flipping mechanism.

When comparing the author's steady-state data with the previous results published by Reddy *et al.*[7] the two studies confirm the same finding on what happens for the second adenine upon enzyme binding in the presence cofactors. In both the author and Reddy *et al.*'s work, EcoP15I flips the second adenine base in the recognition

site in the presence of cofactor. However, the disparity occurs when the author found that in the absence of cofactor, EcoP15I-DNA binary complex still showed a significant enhancement in 2AP fluorescence while Reddy *et al*.'s results showed that no further increase in 2AP fluorescence intensity was observed for M.EcoP15I-DNA complex compared with free duplex for the second 2AP. Furthermore, a more striking difference is what happened for the first adenine base. Reddy *et al*.'s data clearly implied that the first adenine based replaced by 2AP can be flipped out to almost the same extent as for the second adenine base in the recognition site when both M.EcoP15I and cofactors are present. The author's data disagrees with this by showing there was no substantial intensity change for the EcoP15I-DNA binary and EcoP15I-DNA-cofactor ternary complex for the first adenine in the recognition site. The author is not able to repeat Reddy *et al*.'s experiments.

When examining the role of cofactor in terms of base flipping, the author's results shows that irrespective of the presence of SAM or SAH, EcoP15I would flip the target adenine (2^{nd} adenine base within the recognition site) base. Though the interplay amongst cofactors, enzyme and substrate is complex, the author's finding is not unique for EcoP15I as it can also been seen in some other methyltransferases. Lenz *et al.*[31] reported that the lifetime decay parameters of M.TaqI binary and ternary complexes in solution to show the effects of binding with the cofactor analogue, 5'-[2-(amino)ethylthio]-5'-deoxy-adenosine (AETA). The values of lifetimes were almost identical, while there was a difference in the fractional amplitudes. This suggests that essentially similar conformations exist in both complexes, but binding of AETA changes the relative population of conformational states. In addition, Holz *et al.*[27] found that the M.HhaI-DNA binary complex gave comparable steady-state fluorescence intensity to the M.HhaI-DNA-SAH ternary complex. These results both indicate that certain kinds of methyltransferase enzymes are able to perform a base flipping function even without the cofactor. Also, there is some indirect evidence that base flipping also happens in the binary M.HhaI-DNA complex as well. Wu *et al.*[32] showed M.HhaI is capable of catalysing exchange of the proton at the 5-position of cytosine in the absence of cofactor. Base flipping in the binary complex is also supported by the observation that M.HhaI binds more tightly

to a DNA duplex containing thymine, uracil[33] and even adenine or guanine instead of cytosine at target site.[34]

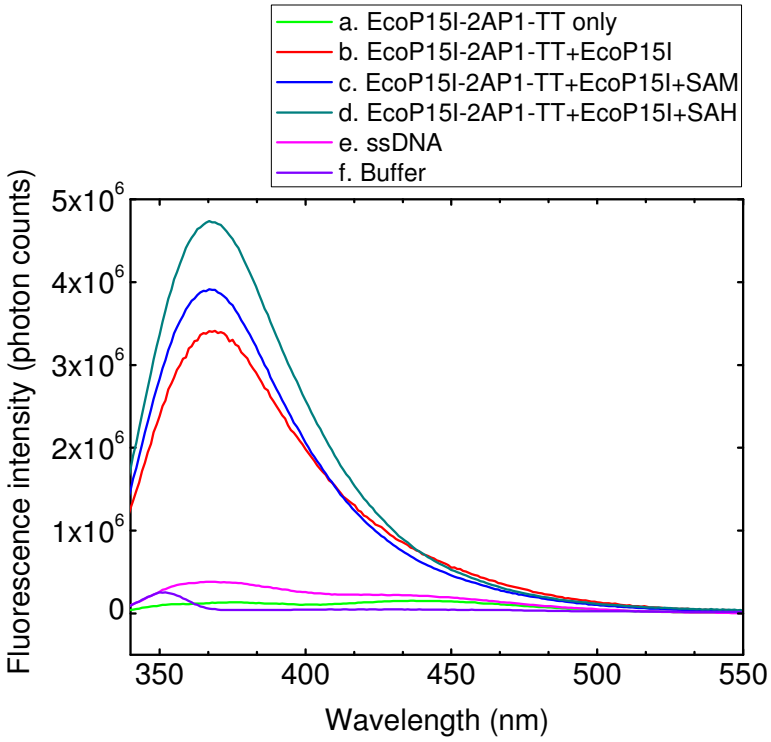


Figure 5.6—Steady-state fluorescence measurements of EcoP15I-2AP1-TT duplex and corresponding ssDNA, EcoP15I-DNA binary and EcoP15I-DNA-SAM/SAH ternary complexes. In the spectra (a)-(e), the buffer background has been subtracted. (f) Emission spectrum of the buffer used in the experiment, shows a typical weak Raman scattering at ~370 nm.

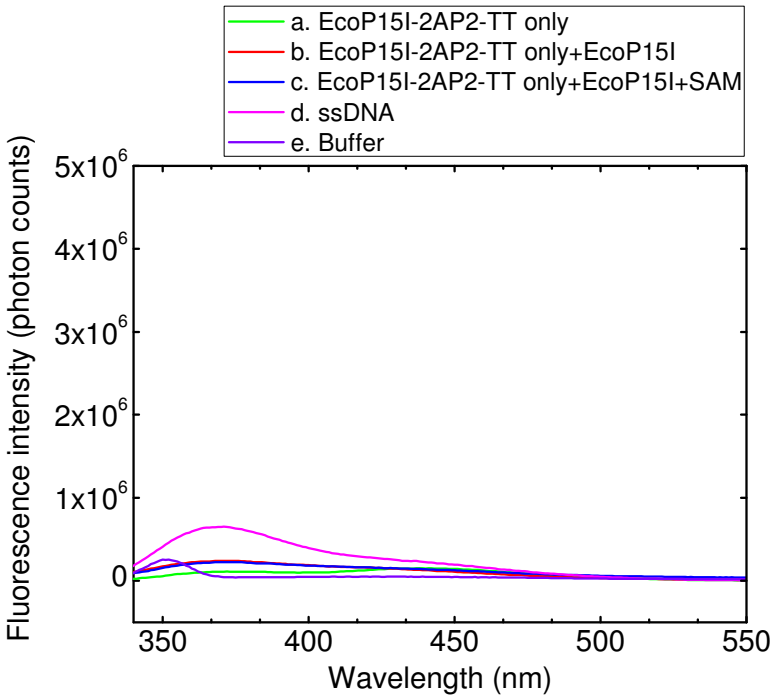


Figure 5.7—Steady-state fluorescence measurements of EcoP15I-2AP2-TT duplex and corresponding ssDNA, EcoP15I-DNA binary and EcoP15I-DNA-SAM ternary complexes. In the spectra (a)-(d), the buffer background has been subtracted. (e) Emission spectrum of the buffer used in the experiment, shows a typical weak Raman scattering at ~370 nm.

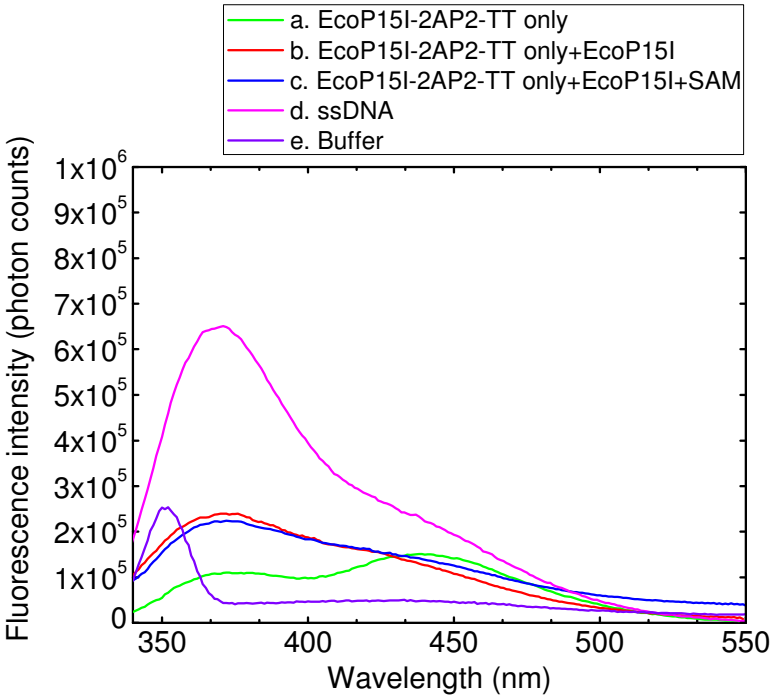


Figure 5.8—close-up of Figure 5.7.

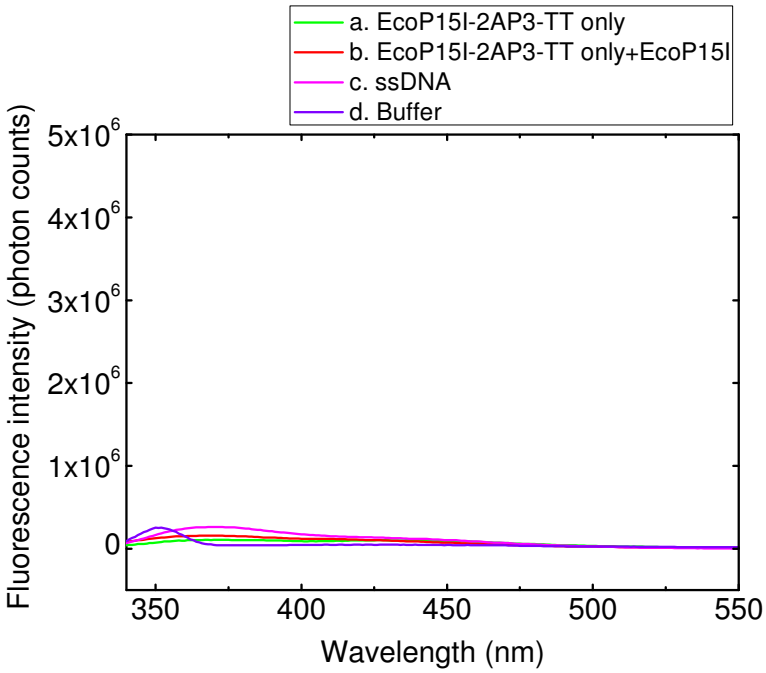


Figure 5.9—Steady-state fluorescence measurements of EcoP15I-2AP3-TT duplex and corresponding ssDNA, EcoP15I-DNA binary complex. In the spectra (a)-(c), the buffer background has been subtracted. (d) Emission spectrum of the buffer used in the experiment, shows a typical weak Raman scattering at ~370 nm.

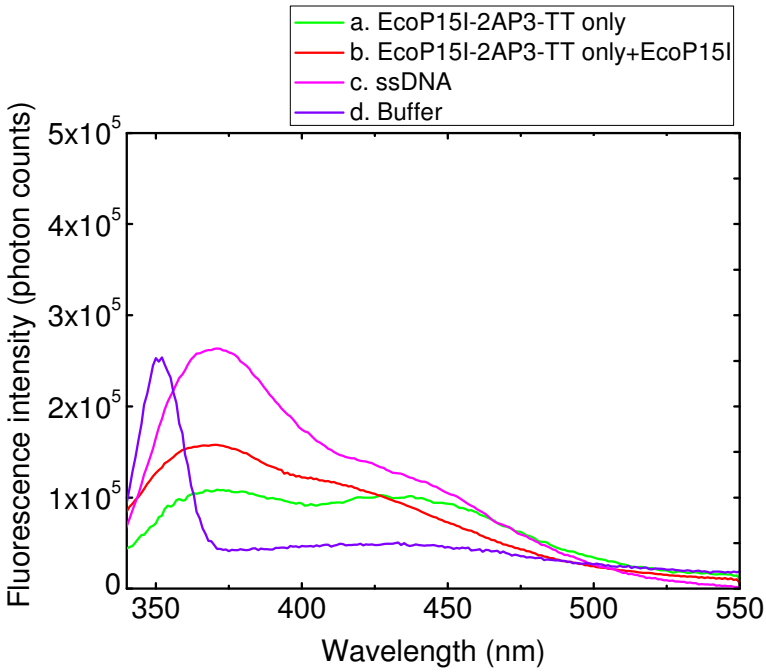


Figure 5.10—A close-up of Figure 5.9.

5.3.3 Time-resolved fluorescence measurements

5.3.3.1 Free duplexes

The lifetime parameters of the three free duplexes (EcoP15I-2AP1-TT, EcoP15I-2AP2-TT and EcoP15I-2AP3–TT duplexes) are essentially the same with minor deviation as shown in Table 5.4. The time-resolved measurements of free duplex revealed four lifetime components. Lifetimes for all three duplexes are all approximately ~50 ps, ~0.5 ns, ~3.0 ns and 10 ns, and fractional amplitude of each is roughly ~75%, ~10%, ~10%, ~5%, which would be due to the very similar sequence context. For each of the duplexes, 2AP is base-paired with thymine and in the middle of cytosine and guanine though with different bases located further from the 2AP. These lifetimes mentioned above all fall in the range of widely accepted four lifetime composition given by 2AP-labeled dsDNA samples, which are <100 ps, ~0.5 ns, ~2 ns, ~10 ns. Also the amplitude of an individual lifetime follows the tendency that the A factor decreases as lifetime increases as previously published[35-37].

It is generally believed that the shortest lifetime corresponds to an intrahelical base state where the fast electron transfer quenching mechanism is dominant, while the longest lifetime is produced by the extrahelical state, in which the 2AP base is free from interbase quenching.[38,39] The observation of four lifetimes can be ascribed to 2AP in the EcoP15I-2AP1-TT duplex existing in four distinct molecular microenvironments thus showing four different quenching rates.

Taking a closer look, for instance, the 2AP base of EcoP15I-2AP1-TT is immediately next to 3' guanine base and also in very close vicinity to another guanine in the opposite strand, which makes τ_1 lifetime of the duplex very short at 40 ps. This is due to the fact that guanine is the most efficient quencher (amongst the four natural DNA bases) of 2AP. A very substantial population (73%) of duplex was in intrahelical state while only 5% of the population was in the extrahelical state, indicative of a very well stacked dsDNA conformation. This situation also applies to the other two duplexes.

5.3.3.2 Free EcoP15I-2AP1-TT duplex, its binary and ternary complexes

As illustrated graphically in Figure 5.11, there are significant differences between the decay parameters of the bound and free duplexes. The decay parameters of the binary (EcoP15I-DNA duplex binary complex) and other two ternary complexes (EcoP15I-DNA duplex-SAM/SAH ternary complex) are nearly identical. Around only 30% of the binary and ternary complex population shows a short ~110 ps lifetime, characteristic of highly stacked 2AP. In contrast, for the free DNA duplex, this value has a much higher fractional amplitude (A_1=0.73). Additionally, the shortest lifetime for the binary and other two ternary complexes (~110 ps) is almost three times longer than the free duplex (~40 ps). Another feature takes place in τ_4, for free DNA duplex, only 5% of the population has the τ_4 lifetime (9.4 ns); while for the binary and ternary complexes, they have a long lifetime ~10.5 ns with A_4 climbing to ~40%. This means the extrahelical conformation becomes dominant for the binary and ternary complexes. The other two intermediate lifetime components are also significantly lengthened for binary and ternary complexes than free dsDNA. τ_2 and τ_3 for free duplex are 0.5 ns 3.0 ns respectively and change to ~1.0 ns and ~5.0 ns for the binary and ternary complexes. The changes in decay parameters of EcoP15I-2AP1-TT duplex on binding of EcoP15I are clearly indicative of the perturbation of intact double helical structure of DNA duplex and eversion of 2AP from a highly stacked intrahelical conformation to an extrahelical conformation.

Previous time-resolved fluorescence studies of methyltransferases have shown the base-flipped complexes to present multiple conformational states, and 2AP base is actually very dynamic in equilibrium among these conformations. The change in 2AP base conformation in this investigation is reflected mostly by the shortest and longest lifetimes. The fractional amplitudes shift is a convincing sign for base flipping and can be observed and verified by the work done within our research group and some others.[31,40-42] Furthermore, the longest lifetime data shows that the flipped 2AP base was located in an environment that was almost identical for binary and ternary complex and where its fluorescence lifetime was similar to that of the free 2AP-ribonucleoside (~10.6 ns). This suggests a total extrahelical base

conformation exposed to a water-like environment and it does not interact evidently with the apolar enzymatic residues in the active site of enzyme, which would shorten its lifetime. This evidence leads the author to think that the binding pocket of EcoP15I is large and filled with water or that the flipped base does not reside within a binding pocket but lies on the exterior of the enzyme.

The lack of EcoP15I-DNA co-crystal structure makes the author unable to judge the contact of 2AP base with the amino acid residues in the enzyme catalysis pocket. Previously, it has been found that due to the direct interaction with 2AP, some aromatic amino acids including tyrosine, phenylalanine and tryptophan can efficiently quench the 2AP base flipped out into the active site of the R-M enzymes.[31,42,43] Fluorescence of 2AP flipped into the active site of methyltransferase TaqI can be efficiently quenched by tyrosine and phenylalanine residues. This gives rise to a subnanosecond decay. The quenching of 2AP fluorescence by tryptophan is known to be rapid, which occurs on the picoseconds timescale, at a rate similar to (or possibly faster than) electron transfer quenching of intrahelical 2AP by guanine.[44] As illustrated before, the flipped 2AP is probably located in a location free from aploar amino acid residue contact. If so the abovementioned aromatic amino acids seem not to be in close proximity to the flipped 2AP base, hence the changes in lifetime and amplitude are caused by the conformational change of 2AP base.

5.3.3.3 Free EcoP15I-2AP2-TT and EcoP15I-2AP3-TT duplexes, and their corresponding binary and ternary complexes

The control duplex of EcoP15I-2AP3-TT and its binary complex give identical lifetime decay parameters, which is consistent with the fact that EcoP15 is unable to flip the 2AP outside the recognition site. Also EcoP15I-2AP2-TT free duplex shows similar decay parameters with its binary and ternary complexes as shown in Figure 5.13. All these facts indicate that EcoP15I doesn't flip the 2AP (first adenine in its recognition sequence) in EcoP15I-2AP2-TT dsDNA. The slight changes in decay parameters come from minor base disturbance upon enzyme binding.

Sample solution	τ_1	τ_2	τ_3	τ_4	A_1	A_2	A_3	A_4
EcoP15I-2AP1-TT only	0.04	0.50	3.0	9.4	0.73	0.12	0.10	0.05
EcoP15I-2AP1-TT + EcoP15I	0.13	1.20	5.5	10.4	0.31	0.10	0.16	0.43
EcoP15I-2AP1-TT + EcoP15I +SAM	0.11	0.80	4.6	10.7	0.36	0.17	0.14	0.34
EcoP15I-2AP1-TT+ EcoP15I + SAH	0.11	0.90	5.1	10.7	0.28	0.08	0.15	0.49
EcoP15I-2AP2-TT only	0.04	0.47	3.1	11.0	0.70	0.12	0.12	0.06
EcoP15I-2AP2-TT + EcoP15I	0.06	0.53	3.2	8.9	0.72	0.13	0.09	0.06
EcoP15I-2AP2-TT + EcoP15I SAM	0.06	0.48	2.7	8.6	0.61	0.22	0.10	0.07
EcoP15I-2AP3-TT only	0.03	0.41	2.7	9.6	0.84	0.08	0.05	0.02
EcoP15I-2AP3-TT + EcoP15I	0.04	0.45	3.1	9.0	0.76	0.11	0.09	0.04

Table 5.4—Fluorescence lifetimes and corresponding fractional amplitudes for free DNA duplexes, EcoP15I-DNA binary complexes, and EcoP15I-DNA-SAM/SAH ternary complexes in aqueous solution.

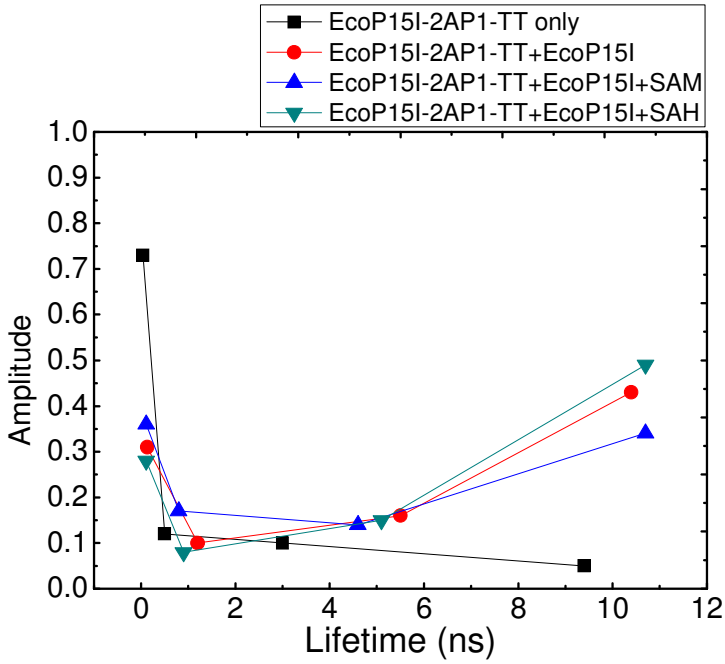


Figure 5.11—Plots of A-factor *vs.* lifetime for EcoP15I-2AP1-TT free duplex, its corresponding binary and ternary complexes.

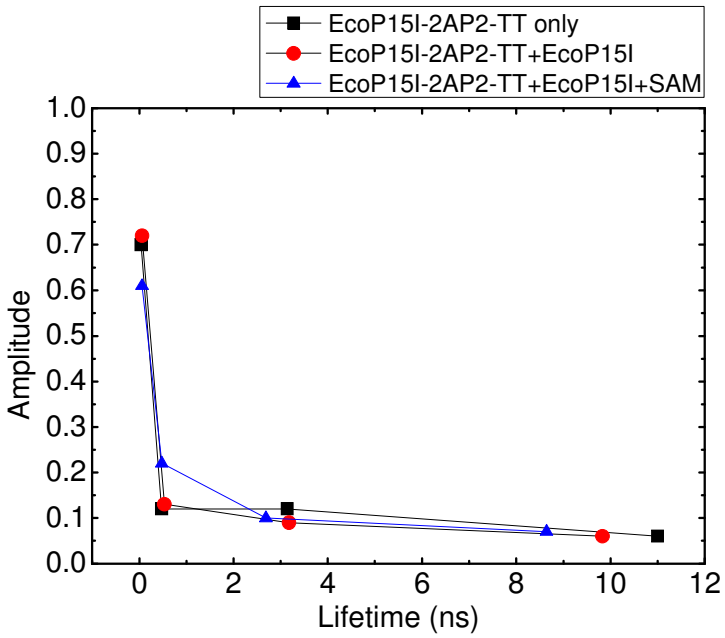


Figure 5.12— Plots of A-factor *vs.* lifetime for EcoP15I-2AP2-TT free duplex, its corresponding binary and ternary complexes.

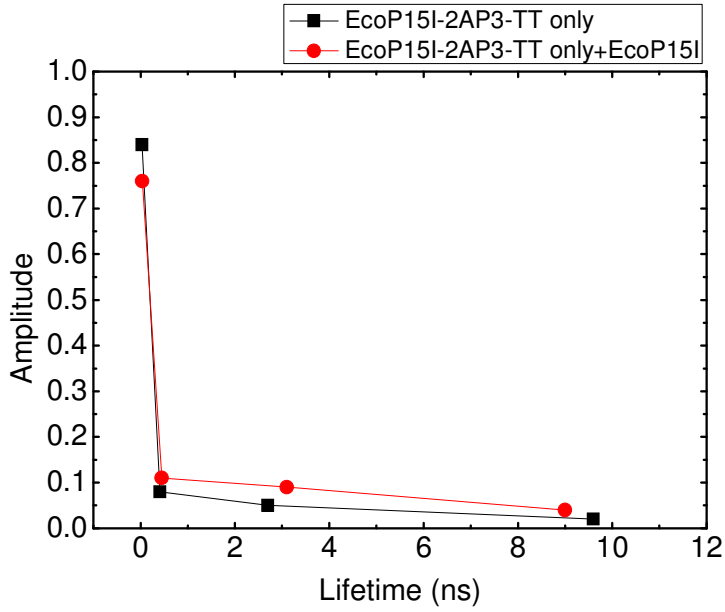


Figure 5.13—Plots of A-factor *vs.* lifetime for EcoP15I-2AP3-TT free duplex, its corresponding binary and ternary complexes.

5.3.4 Cleavage of matched and mismatched DNA samples by EcoP15I

DNA substrates that contain two adjacent recognition sites can be cleaved efficiently by EcoP15I as illustrated before[5]. However, little has been known about what will happen if one of the two recognition site contains a mismatch or even one of the 'recognition site' is changed to 5'-CTGCTG-3' rather than 5'-CAGCAG-3'.

The author varied the base composition of two adjacent recognition sites in several 120 mer DNA duplexes. Two reversely orientated CAGCAG with a head to head arrangement formed in duplex 1, as shown in Table 5.4. The other duplexes varied the composition of these two recognition sites, with some of them have one 5'-CAGCAG-3' site and another 5'-CAGCTG-3', 5'-CTGCAG-3' or 5'-CTGCTG-3' site on the opposite strand. A or T base in between the C and G base is base-paired either with its nature complementary strand or form A:A or T:T mismatch. Nine 120 mer DNA duplexes were used for conducting this experiment. The reaction of free duplexes and duplexes incubated with EcoP15I were all analysed on a 15% polyacrylamide gel. The results are shown in Figure 5.14, three duplexes can be cut by EcoP15I in lane 2, 6, 16. Clearly two additional bands corresponding to DNA length shorter than 100 bp can be seen in these lanes.

The reason for the author to find out the cleavage pattern of EcoP15I is that this enzyme can be utilised for exploring DNA containing trinucleotides CAG repeats, which can be found in various neurological disorders.[9,45] CAG trinucleotide repeats in the coding region of human genes can cause a neurodegenerative disorder when the length of a repeated section of a gene surpasses normal range. For instance, the *HTT* gene located on the short arm of chromosome 4 is composed of 67 exons. The first exon contains the disease-generating CAG repeats expansion (i.e. ... CAGCAGCAG ...). CAG can be translated into amino acid glutamine, thus the expansion of them results in the generation of a chain of glutamine known as a polyglutamine stretch (or polyQ stretch) in expressed protein. This polyglutamine stretch renders the protein neurotoxic, which gives rise to HD.[10] The threshold of the

number for CAG repeats for giving rise to HD is ~36, which means people carrying CAG tract length longer than 36 in HD gene are prone to be affected.[46] There is an approximate correlation between the age of onset of the disease and the length of the CAG tract as well. Super-long CAG expansions are found in brains of HD patients[47] and it can also be introduced into R6/2 transgenic mice[48].

The three duplexes which can be cut by EcoP15I are all listed in Figure 5.15 and by comparing the sequence composition of those and others, a model in terms of two adjacent recognition sites for efficient cleavage can be revealed. As shown in Figure 5.15, EcoP15I requires two 5'-CAGCAG-3' sequences on the opposite strand of dsDNA. One recognition site must be fully matched while the other CAGCAG site allows the first and second adenine base pair with thymine or adenine. Our finding could be potentially used for sizing CAG repeats. This requires the genomic DNA encompassing an inverse Ecop15I recognition site; alternatively this inverse recognition site could be introduced by PCR for the DNA sample subject to analysis.

However, as shown in Figure 5.16, DNA contains super-long CAG or CTG repeats has a strong tendency to form unusual structure creating 3WJ. The 'protruding' hairpin loop of 3WJ usually contain A:A or T:T mismatched bases.[49,50] Thereby if one expects to apply EcoP15I on investigating DNA containing CAG or CTG expansions, further work is required to be done to see if EcoP15I is able to cut DNA 3WJ containing A:A mismatch in the hairpin region.

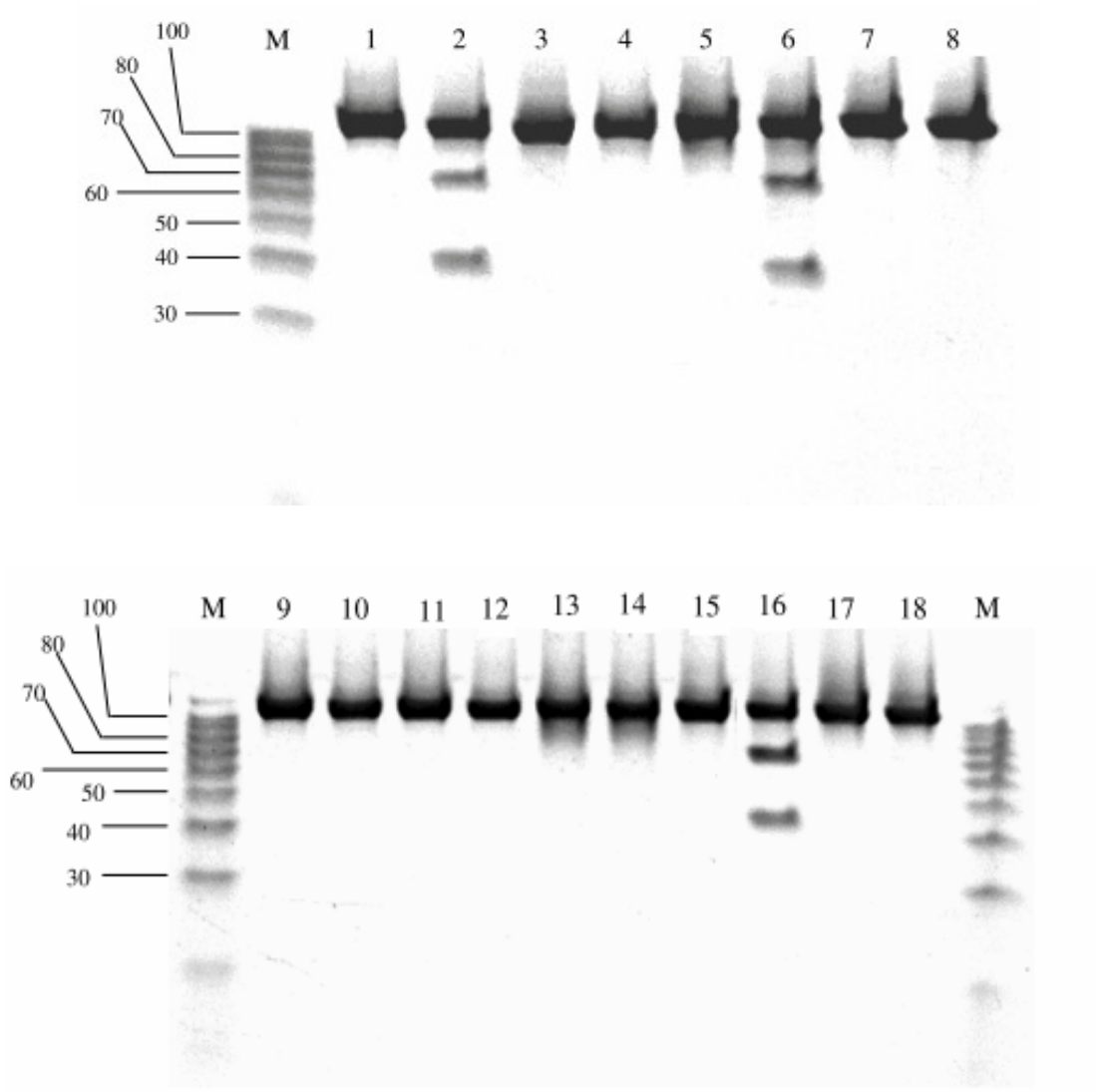


Figure 5.14—EcoP15I matched and mismatched cutting revealed by 15% PAGE. The samples were loaded into the gel as follows: (1) Duplex1 only. (2) Duplex1 + EcoP15I. (3) Duplex2 only. (4) Duplex2 + EcoP15I. (5) Duplex3 only. (6) Duplex3 + EcoP15I. (7) Duplex4 only. (8) Duplex4 + EcoP15I. (9) Duplex5 only. (10) Duplex5 + EcoP15I. (11) Duplex6 only. (12) Duplex6 + EcoP15I. (13) Duplex7 only. (14) Duplex7 + EcoP15I. (15) Duplex8 only. (16) Duplex8 + EcoP15I. (17) Duplex9 only. (18) Duplex9 + EcoP15I. The sequences of all nine duplex were shown in Table 5.4 above.

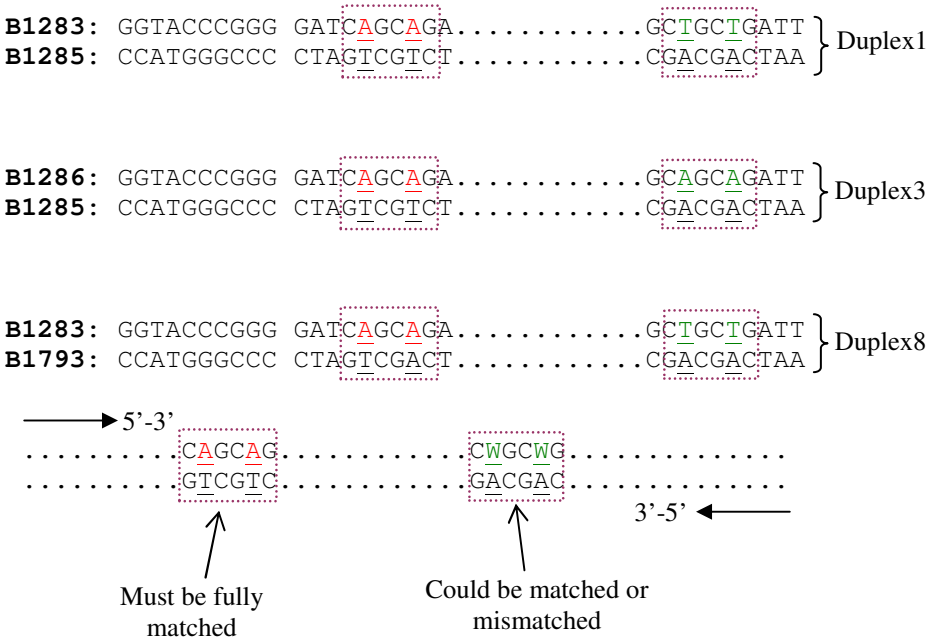


Figure 5.15—The diagram which shows requirement of two adjacent recognition sites on dsDNA for efficient EcoP15I cleavage deduced from the EcoP15I dsDNA cleavage experiments.

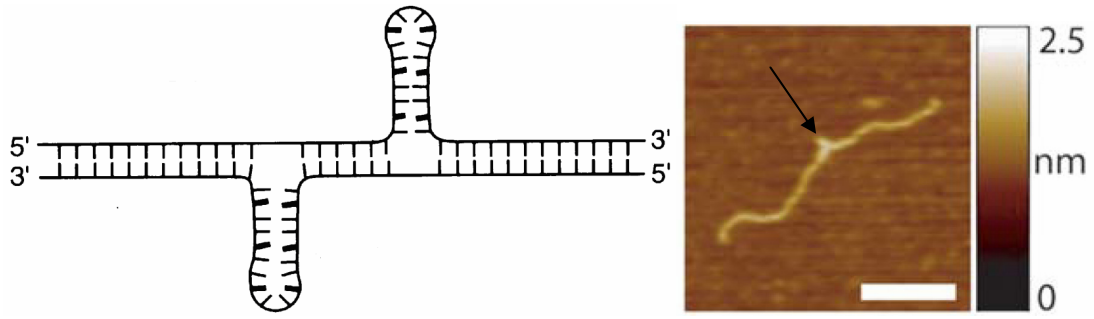


Figure 5.16—Unusual DNA structure caused by super-long CAG repeats creating 3WJ. (Left): an schematic diagram of 3WJ DNA structure caused by CAG repeats (reprinted from Sinden *et al.*[51]). (Right) 'protruding' 3WJ DNA structure (pointed by the arrow) caused by CAG repeats revealed by AFM (reprinted from Duzdevich *et al.*[52]).

5.4 References

1. M. R. Tock and D. T. F. Dryden, *Curr. Opin. Microbiol.*, 2005, 8, 466-472.
2. T. A. Bickle and D. H. Krüger, *Microbiol. Rev.*. 1993, 57, 434-450.
3. P. Bist, S. Sistla, V. Krishnamurthy, A. Acharya, B. Chandrakala and D. N. Rao, *J. Mol. Biol.*, 2001, 310, 93-109.
4. M. Mücke, S. Reich, E. Möncke-Buchner, M. Reuter and D. H. Krüger, *J. Mol. Biol.*, 2001, 312, 687-698.
5. P. Janscak, U. Sandmeier, M. D. Szczelkun and T. A. Bickle, *J. Mol. Biol.*, 2001, 306, 417-431.
6. K. Wagenführ, S. Pieper, P. Mackeldanz, M. Linscheid, D. H. Krüger and M. Reuter, *J. Mol. Biol.*, 2007, 366, 93-102.
7. Y. V. R. Reddy and D. N. Rao, *J. Mol. Biol.*, 2000, 298, 597-610.
8. E. Moencke-Buchner, S. Reich, M. Muecke, M. Reuter, W. Messer, E. E. Wanker and D. H. Krueger, *Nucleic Acids Res.*, 2002, 30, e83/81-e83/87.
9. J. F. Gusella and M. E. MacDonald, *Nat. Rev. Neurosci.*, 2000, 1, 109-115.
10. H. T. Orr and H. Y. Zoghbi, *Annu. Rev. Neurosci.*, 2007, 30, 575-621.
11. C. E. Pearson and R. R. Sinden, *Curr. Opin. Struct. Biol.*, 1998, 8, 321-330.
12. S. M. Mirkin, *Nature*, 2007, 447, 932-940.
13. E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel and A. Bairoch, 2005.
14. E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel and A. Bairoch, *Nucleic Acids Res.*, 2003, 31, 3784-3788.
15. R. Owczarzy, A. V. Tataurov, Y. Wu, J. A. Manthey, K. A. McQuisten, H. G. Almabrazi, K. F. Pedersen, Y. Lin, J. Garretson, N. O. McEntaggart, C. A. Sailor, R. B. Dawson and A. S. Peek, *Nucleic Acids Res.*, 2008, 36, W163-W169.
16. D. T. F. Dryden, L. P. Cooper, P. H. Thorpe and O. Byron, *Biochemistry*, 1997, 36, 1065-1076.
17. G. A. Roberts, L. P. Cooper, J. H. White, T.-J. Su, J. T. Zipprich, P. Geary, C. Kennedy and D. T. F. Dryden, *Nucleic Acids Res.*, 2011.
18. T. J. Su, B. A. Connolly, C. Darlington, R. Mallin and D. T. F. Dryden,

Nucleic Acids Res., 2004, 32, 2223-2230.

19. E. Y. M. Bonnist, K. Liebert, D. T. F. Dryden, A. Jeltsch and A. C. Jones, *Biophys. Chem.*, 2012, 160, 28-34.
20. I. Ahmad and D. N. Rao, *J. Mol. Biol.*, 1994, 242, 378-388.
21. K. H. Wyszomirski, U. Curth, J. Alves, P. Mackeldanz, E. Möncke-Buchner, M. Schutkowski, D. H. Krüger and M. Reuter, *Nucleic Acids Res.*, 2011.
22. M. Rist, H.-A. Wagenknecht and T. Fiebig, *ChemPhysChem*, 2002, 3, 704-707.
23. E. Y. M. Bonnist and A. C. Jones, *ChemPhysChem*, 2008, 9, 1121-1129.
24. T.-J. Su, M. R. Tock, S. U. Egelhaaf, W. C. K. Poon and D. T. F. Dryden, *Nucleic Acids Res.*, 2005, 33, 3235-3244.
25. X. Cheng and R. J. Roberts, *Nucleic Acids Res.*, 2001, 29, 3784-3795.
26. B. W. Allan and N. O. Reich, *Biochemistry*, 1996, 35, 14757-14762.
27. B. Holz, E. Weinhold, S. Klimasauskas and S. Serva, *Nucleic Acids Res.*, 1998, 26, 1076-1083.
28. E. G. Malygin, A. A. Evdokimov, V. V. Zinoviev, L. G. Ovechkina, W. M. Lindstrom, N. O. Reich, S. L. Schlagman and S. Hattman, *Nucleic Acids Res.*, 2001, 29, 2361-2369.
29. S. S. Szegedi, N. O. Reich and R. I. Gumport, *Nucleic Acids Res.*, 2000, 28, 3962-3971.
30. B. W. Allan, J. M. Beechem, W. M. Lindstrom and N. O. Reich, *J. Biol. Chem.*, 1998, 273, 2368-2373.
31. T. Lenz, E. Y. M. Bonnist, G. Pljevaljčić, R. K. Neely, D. T. F. Dryden, A. J. Scheidig, A. C. Jones and E. Weinhold, *J. Am. Chem. Soc.*, 2007, 129, 6240-6248.
32. J. C. Wu and D. V. Santi, *J. Biol. Chem.*, 1987, 262, 4778-4786.
33. A. S. Yang, J.-C. Shen, J.-M. Zingg, S. Mi and P. A. Jones, *Nucleic Acids Res.,* 1995, 23, 1380-1387.
34. S. Klimasauskas and R. J. Roberts, *Nucleic Acids Res.*, 1995, 23, 1388-1395.
35. O. J. G. Somsen, L. B. Keukens, M. N. de Keijzer, A. van Hoek and H. van Amerongen, *ChemPhysChem*, 2005, 6, 1622-1627.
36. T. M. Nordlund, S. Andersson, L. Nilsson, R. Rigler, A. Graeslund and L. W.

McLaughlin, *Biochemistry*, 1989, 28, 9095-9103.

37. C. R. Guest, R. A. Hochstrasser, L. C. Sowers and D. P. Millar, *Biochemistry*, 1991, 30, 3271-3279.
38. E. L. Rachofsky, E. Seibert, J. T. Stivers, R. Osman and J. B. A. Ross, *Biochemistry*, 2001, 40, 957-967.
39. R. A. Hochstrasser, T. E. Carver, L. C. Sowers and D. P. Millar, *Biochemistry*, 1994, 33, 11971-11979.
40. R. K. Neely, D. Daujotyte, S. Grazulis, S. W. Magennis, D. T. F. Dryden, S. Klimasauskas and A. C. Jones, *Nucleic Acids Res.*, 2005, 33, 6953-6960.
41. B. Youngblood, E. Bonnist, D. T. F. Dryden, A. C. Jones and N. O. Reich, *Nucleic Acids Res.,* 2008, 36, 2917-2925.
42. R. K. Neely, G. Tamulaitis, K. Chen, M. Kubala, V. Siksnys and A. C. Jones, *Nucleic Acids Res.*, 2009, 37, 6859-6870.
43. T. Xia, H.-C. Becker, C. Wan, A. Frankel, R. W. Roberts and A. H. Zewail, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, 100, 8119-8123.
44. T. Fiebig, C. Wan and A. H. Zewail, *ChemPhysChem*, 2002, 3, 781-788.
45. K. Usdin and E. Grabczyk, *Cell. Mol. Life Sci.*, 2000, 57, 914-931.
46. N. S. Wexler, J. Lorimer, J. Porter, F. Gomez, C. Moskowitz, E. Shackell, K. Marder, G. Penchaszadeh, S. A. Roberts, J. Gayan, D. Brocklebank, S. S. Cherny, L. R. Cardon, J. Gray, S. R. Dlouhy, S. Wiktorski, M. E. Hodes, P. M. Conneally, J. B. Penney, J. Gusella, J.-H. Cha, M. Irizarry, D. Rosas, S. Hersch, Z. Hollingsworth, M. MacDonald, A. B. Young, J. M. Andresen, D. E. Housman, Y. M. M. De, E. Bonilla, T. Stillings, A. Negrette, S. R. Snodgrass, M. D. Martinez-Jaurrieta, M. A. Ramos-Arroyo, J. Bickham, J. S. Ramos, F. Marshall, I. Shoulson, G. J. Rey, A. Feigin, N. Arnheim, A. Acevedo-Cruz, L. Acosta, J. Alvir, K. Fischbeck, L. M. Thompson, A. Young, L. Dure, C. J. O'Brien, J. Paulsen, A. Brickman, D. Krch, S. Peery, P. Hogarth, D. S. Higgins, Jr. and B. Landwehrmeyeri, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, 101, 3498-3503.
47. M. DiFiglia, E. Sapp, K. O. Chase, S. W. Davies, G. P. Bates, J. P. Vonsattel and N. Aronin, *Science*, 1997, 277, 1990-1993.
48. Z. DeMarch, C. Giampa, S. Patassini, G. Bernardi and F. R. Fusco, *Neurobiol.*

Dis., 2008, 30, 375-387.

49. M. Mitas, *Nucleic Acids Res*., 1997, 25, 2245-2253.
50. C. E. Pearson, Y.-H. Wang, J. D. Griffith and R. R. Sinden, *Nucleic Acids Res*., 1998, 26, 816-823.
51. R. R. Sinden, V. N. Potaman, E. A. Oussatcheva, C. E. Pearson, Y. L. Lyubchenko and L. S. Shlyakhtenko, *J. Biosci*., 2002, 27, 53-65.
52. D. Duzdevich, J. Li, J. Whang, H. Takahashi, K. Takeyasu, D. T. F. Dryden, A. J. Morton and J. M. Edwardson, *PLoS ONE*, 2011, 6, e17119.

Chapter 6---Biophysical Characterisation of Type II Restriction Enzyme TseI Revealing That it Cuts A:A and T:T Mismatches in CAG and CTG Repeats

6.1 Introduction

6.1.1 Background

Type II restriction endonucleases have revolutionised molecular biology because of their ability to cleave DNA molecules at defined locations within or near to a specific base pair sequence.[1-3] These kinds of enzymes have been extensively used in many areas comprising recombinant DNA technology, genome mapping, and genetic manipulation[4]. Thousands of restriction enzymes are known, and many are commercially available (REBASE[5,6]: http://rebase.neb.com/rebase/rebase.html). The TseI[7] restriction enzyme isolated from a thermophilic bacterium shows optimal activity at an elevated temperature around 65°C. It recognises sequence 5' GCWGC-3' / 3'-CGWCG-3' (where W=A or T) and cleaves after the first G on the top strand and last C on the bottom strand to produce fragments with three base 5' overhangs, leaving 3'-hydroxyls and 5'-phosphate ends.[1]

Given its recognition sequence, TseI cleaves CAG and CTG trinucleotide repeats, which are also involved in the aetiology of a number of neurodegenerative diseases such as HD (CAG repeats). Previous reporters have shown that the various triplet repeats give rise to unusual non-B form DNA structures, including triplexes, hairpins, slipped-strand DNA, and G-quadruplexes.[8] Recently, AFM was used to analyse DNA samples of various CAG repeat lengths.[9] It has been found that the structural profile of the DNA changed significantly as CAG repeat length increased. DNA from wild type mice having low CAG repeat numbers appeared as short linear molecules, whereas when the CAG repeat length increased, various DNA structures, including convolutions, folds and protrusions, became apparent. Over half of DNA molecules having 408 CAG repeats molecules showed one of these unusual structures. It was demonstrated that the convoluted DNA was subject to mung bean nuclease cleavage, indicating that it contained hairpin mismatches which have single-stranded regions in the DNA. TseI was used to further characterise the structures observed in the super-long CAG repeats. It was found that at room temperature, TseI preferentially

cleaved linear regions of 858 CAG repeats DNA, leaving behind the contorted regions. In contrast, at 80°C TseI completely digested the DNA. These observations suggested that TseI prefers to cleave CAG repeats within normal B-form DNA, but that at higher temperatures it is also able to cleave DNA containing A:A or T:T mismatches at the centre of the recognition sequence (5'-GC(A/T)GC-3'), which will be present in the hairpins as shown in Figure 6.1.

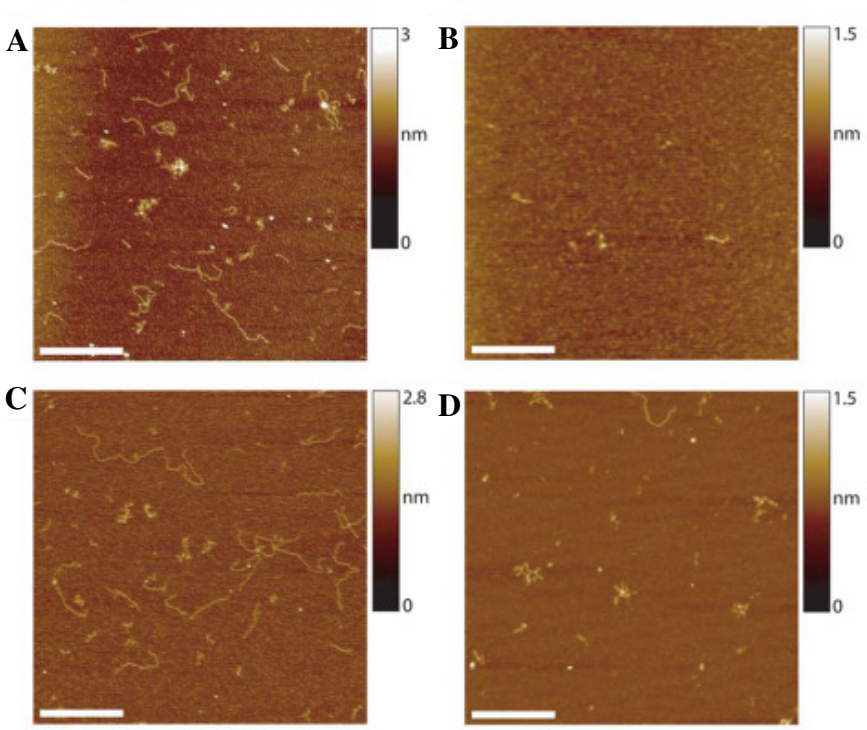


Figure 6.1—Super-long CAG containing DNA before and after TseI digestion (reprinted from Duzdevich *et al.*[9]). (A) DNA containing 585 CAG repeats in the absence of TseI digestion, some of which shows a linear structure; while others are highly folded or convoluted. (B) DNA containing 585 CAG repeats after extensive Tse1 digestion at 80℃. A dense 'background' resulting from short 6 bp fragments remains together with a few aggregations of these fragments. (C) DNA containing 585 CAG repeats without digestion. (D) DNA containing 560 CAG repeats after TseI digestion at ambient temperature. The result indicates that TseI enzyme has preferentially cleaved linear regions of DNA, leaving behind the anomalous regions.

The author speculates that TseI may belong to the group of restriction enzymes which flip out nucleotides within a recognition sequence as described by Bochtler *et al.*[10] and further expanded by Neely *et al.*[11] Restriction endonuclease Ecl18kI recognises the sequence /CCNGG and cleaves it before the outer cytosine thereby generating 5 nt 5'-overhangs. It has been suggested that Ecl18kI is highly related to NgoMIV in evolution, which cleaves the sequence G/CCGGC and leaves 4 nt 5'-overhangs. By crystallising the Ecl18kI-DNA complex, Bochtler *et al.*[10] found that Ecl18kI flips both central nucleotides within the CCNGG sequence and buries the extruded bases in pockets within the protein. Nucleotide flipping results in disruption of Watson-Crick base pairing producing a kink in the DNA and shifts the duplex register by 1 bp, making the distances between scissile phosphates in Ecl18kI and NgoMIV co-crystal structures almost identical as shown in Figure 6.2. Therefore, these two enzymes can use a conserved DNA recognition module to recognise different sequences and generate different cleavage patterns. It was suggested that base nucleotide flipping is sufficiently facile to occur in many occasions, and it may be found in other type II restriction enzyme-DNA complexes. Bochtler *et al.*[10] reasoned that in Ecl18kI, nucleotide flipping plays a very different role in order to adjust the cleavage pattern: the flipped-out complex having the central base extruded reduces the length of a 5 bp stretch recognition/cleavage site to a 1 bp shorter canonical 4 bp stretch, which accounts for the differences in the cleavage positions of Ecl18kI (5 nt 5'-overhang CCNGG, N stands for the any nucleotide) and NgoMIV (4 nt 5'-overhang CCGG). Neely *et al.*[11] studied the base flipping mechanism adopted by some restriction enzymes using time-resolved fluorescence of 2AP. They proposed a base flipping process and locking steps prior to DNA cleavage by enzyme based on their lifetime measurement and previous studies[12,13] as shown in Figure 6.3. It was found by them that Ecl18kI uses a two-step mechanism to achieve base flipping: firstly, breaking the DNA duplex and secondly, capturing the extruded base. In contrast to this, PspGI was suggested to use a double-check mechanism, which is to say, target bases are flipped only transiently without being locked, as a result of base recognition by PspGI.

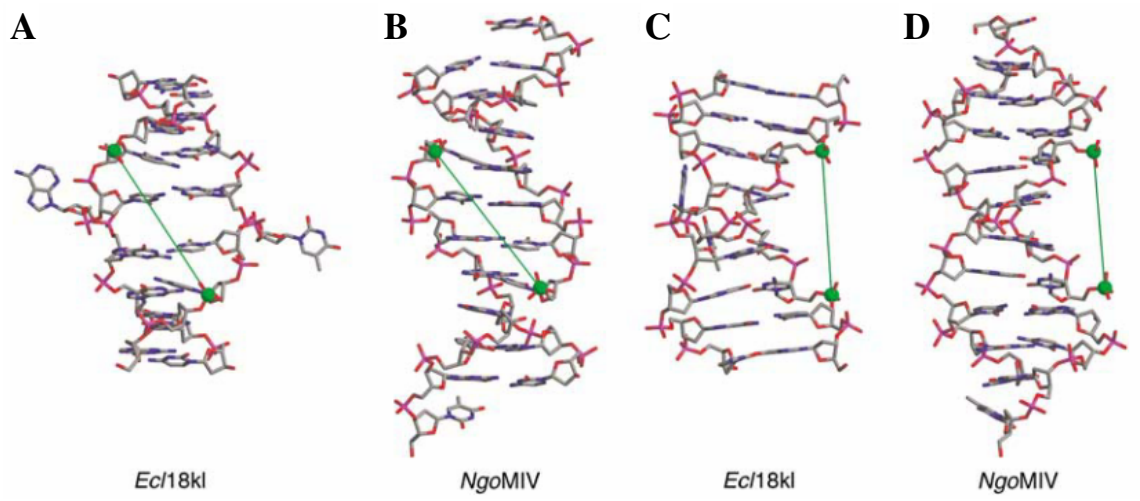


Figure 6.2—Conformation of DNA in EcI18kI-DNA complex (A, C) and in NgoMIV-DNA (B, D) (reprinted from Bochtler *et al.*[10]). (C) and (D) are presented by rotating (A) and (B) by 90℃ respectively around the vertical axis. It is clearly seen that the central nucleotide pair is flipped out of the DNA helix in (A, C). Phosphorus atoms at scissile phosphates are shown in green balls. The distance between these phosphorus atoms is 17.2 Å in both EcI18kI-DNA and NgoMIV-DNA complexes.

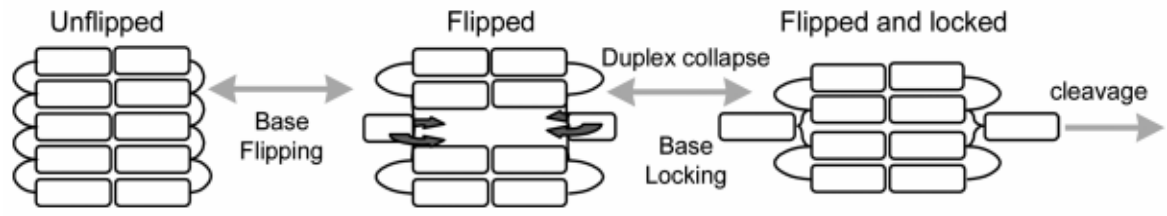


Figure 6.3—Schematic illustrations of the DNA bases (white rectangles) during the flipping process and subsequent locking steps prior to effective DNA cleavage (reprinted from Neely *et al.*[11]). The target base can be in a highly dynamic flipped (but not locked) state (predominant in the PspGI-DNA complex) or locked in the enzyme flipping pocket (as in WT EcI18kI-DNA and EcoRII-C-DNA complexes).

6.1.2 Aims of this chapter

In the few studied cases, the restriction enzyme uses a flipping mechanism to recognize a sequence or adjust the shape of the DNA in order to result in a favourable geometry within its catalytic site for enzymatic cleavage. Either of these

scenarios might occur when TseI encounters cognate dsDNA. If TseI were able to cut an A:A or a T:T central base mismatch in its recognition site, the author thought that nucleotide flipping mechanism may allow TseI to recognize and cleave near an A:A or a T:T mismatch by collapsing the recognition sequence to 5'-GCGC-3', thereby ignoring mismatches. In the current study the author set out to test this idea. The fundamental experiments in enzymology including activity, affinity for its cognate substrate, solution assembly, are also required to be performed since it can contribute to the understanding and application of TseI in certain areas.

6.2 Experimental

6.2.1 TseI restriction enzyme

TseI (5000 units/mL, New England Biolabs) was used as supplied. The gene sequence was deposited in GenBank with accession number CS808196.1 as shown in Table 6.1. This stock sample had a concentration of 10.8 μM in terms of TseI monomers calculated using a molecular weight of 42986.8 Da including the N-terminal methionine (390 amino acid residues in total) and an extinction coefficient at 280 nm of 37930 M^{-1} cm^{-1} assuming all cysteine residues were reduced. The ExPASy-ProtParam tool[14] at http://web.expasy.org/cgi-bin/protparam/protparam was used to calculate these values.

GenBank: CS808196.1
SOURCE: Thermus sp.
ORGANISM: Thermus sp.

```
1   atgaaaagat tagcaggctt aataagctta gcagacttaa tacaaggtga tactgagttt
61  aagataagct gggaaaaccg agggaaaaag gcgctcactc ttctggccga gaaggcaggc
121 atcagatgcg acgagcagct agatgatctt ctgtcgcaag ccctggatct tgcaaggagc
181 acgcttacct ccggcaaaaa tcctgatgct gacatcgctc acttctggga ggaggtcgaa
241 aaaaacgcca ccctcttaac gaaaaacgac tacctccgag cggctgtagt agctctttcg
301 tttgcccacc gctttgcccg aacagactac ggatcgtcaa ggcaacgcgg cttcgggcaa
361 ctctggggag atgcgattca aggcttcctt ggtgaaattg ccttccagaa gtttatgagg
421 tcagccacgt ctgggaggac catccctatt ttagacgcca gcgaagaaga tcttggagtc
481 gccctaagcg ctgacatagt tgaagtcatc acagagggga aatcaataaa gccctcaaaa
541 agaatcagca tcaagactac gaagctccat gggcgctggt tagatgtacc ctacgctcaa
601 aataagcaca gcgacattta cgttctggtt aaagtcggga ctgacgccga tgcgcttttc
661 aactttctgg caagcgtagg ggcgcttgag aaagtcttaa ccgcctatca agagggcggt
721 cttgctgaag gcgagcttcc ttttctcaac gaaggcgaag cgctcaaaag agctaaggaa
781 gaggtagaaa aaatgaagga aaaaaacatg ctttttttag cctttatagc tggctggaag
841 gagaaggatc ggctcagcca aaccttcgaa gctcacgagc acaacgccca aagagcccgc
901 acaaaaatca ctgtctacag cggagttggt acaatttcat ctggtagcgt gcgaacaaag
961 caaatcacct ttcgcggtcc cctccctaaa aacaatctgc tggttgagtt ttatccaata
1021ggaaaattct ccaaaagcca gcatgcactg tgcagcacag atctgcttgt gaaggatctc
1081aataagatag cagaacttct ctctgctcct gaagaggggg atgaatgcgc acagtaa
```

```
METKRLAGLI  SLADLIQGDT  EFKISWENRG  KKALTLLAEK  AGIRCDEQLD  DLLSQALDLA
RSTLTSGKNP  DADIAHFWEE  VEKNATLLTK  NDYLRAAVVA  LSFAHRFART  DYGSSRQRGF
GQLWGDAIQG  FLGEIAFQKF  METRSATSGR  TIPILDASEE  DLGVALSADI  VEVITEGKSI
KPSKRISIKT  TKLHGRWLDV  PYAQNKHSDI  YVLVKVGTDA  DALFNFLASV  GALEKVLTAY
QEGGLAEGEL  PFLNEGEALK  RAKEEVEKME  TKEKNMETLF  LAFIAGWKEK  DRLSQTFEAH
EHNAQRARTK  ITVYSGVGTI  SSGSVRTKQI  TFRGPLPKNN  LLVEFYPIGK  FSKSQHALCS
TDLLVKDLNK  IAELLSAPEE GDECAQSTOP
```

Table 6.1—The nucleotide and amino acid sequence of TseI.

6.2.2 Oligonucleotides used in this chapter

The molecular weight of oligonucleotides was calculated by OligoCalc: Oligonucleotide Properties Calculator[15] (http://www.basic.northwestern.edu/biotools/oligocalc.html); melting temperature (T_m) and extinction coefficient were calculated using Integrated DNA Technologies biophysics online software[16] (http://biophysics.idtdna.com/). Note the T_m was calculated under the 50 nM concentration, 50 mM K^+ and 10 mM Mg^{2+} ion condition. The monoisotopic molecular mass of all the synthesised olionucleotides was calculated using Oligo II Mass Calculator v1.0 (http://library.med.utah.edu/masspec/oligoii.htm) and the molecular formulae of synthesised olionucleotides were calculated using SPARTAN software.

All oligonucleotides used in this chapter were listed in Table 6.2. 2AP/T duplex 1, A/2AP duplex 2, T/2AP duplex 3, with 2AP base either in the centre of the recognition sequence or outside the recognition sequence. All three duplexes were used in the TseI-DNA interaction study, in order to investigate if TseI would be able to flip the central base by steady-state fluorescence measurements. Anisotropy duplex was the one used in the anisotropy measurements upon addition of TseI. Fluorescence assay duplex was the fully matched duplex used in the TseI activity assay for cutting matched substrate, whereas the fluorescence assay mismatched duplex with a A:A mismatch in the centre of recognition site, which was used in the TseI activity assay for cutting matched substrate. A:T, A:A, T:T, G:C, G:G duplexes were used in the gel electrophoresis and denature HPLC for investigating the TseI matched duplex and mismatched duplex cleavage. The other four shorter 12, 13, 15, 16 mer product sequences were used for conducting the HPLC experiments to determine the retention times of cleaved fragments DNA generated in the TseI matched duplex and mismatched duplex cleavage experiments. Note that 15 and 16 mer product sequences both have a phosphate group at the 5'-end.

Duplex name	Sequence
2AP/T duplex 1	AGGAGTGAAGTC**GC2GC**CCGTGCTCAAG (5′–3′) TCCTCACTTCAG**CGTCG**GGCACGAGTTC (3′–5′)
A/2AP duplex 2	AGGAGTGAAGTC**GCAGC**CCGTGCTCAAG (5′–3′) TCCTCACTTCAG**CG2CG**GGCACGAGTTC (3′–5′)
T/2AP duplex 3	AGGAGTGAAGTC**GCAGC**CCGTGCTCAAG (5′–3′) TCCTCACTTC**2**G**CGTCG**GGCACGAGTTC (3′–5′)
2AP/A duplex 4	AGGAGTGAAGTC**GC2GC**CCGTGCTCAAG (5′–3′) TCCTCACTTCAG**CGACG**GGCACGAGTTC (3′–5′)
2AP/2AP duplex 5	AGGAGTGAAGTC**GC2GC**CCGTGCTCAAG (5′–3′) TCCTCACTTCAG**CG2CG**GGCACGAGTTC (3′–5′)
Anisotropy duplex	HEX –AGGAGTGAAGTC**GCAGC**CCGTGCTCAAG (5′–3′) TCCTCACTTCAG**CGTCG**GGCACGAGTTC (3′–5′)
Fluorescence assay duplex	HEX –AGGAGTGAAGTC**GCAGC**CCGTGCTCAAG (5′–3′) BHQ1–TCCTCACTTCAG**CGTCG**GGCACGAGTTC (3′–5′)
Fluorescence assay mismatched duplex	HEX –AGGAGTGAAGTC**GCAGC**CCGTGCTCAAG (5′–3′) BHQ1–TCCTCACTTCAG**CGACG**GGCACGAGTTC (3′–5′)
A/T duplex	AGGAGTGAAGTC**GCAGC**CCGTGCTCAAG (5′–3′) TCCTCACTTCAG**CGTCG**GGCACGAGTTC (3′–5′)
A/A duplex	AGGAGTGAAGTC**GCAGC**CCGTGCTCAAG (5′–3′) TCCTCACTTCAG**CGACG**GGCACGAGTTC (3′–5′)
T/T duplex	AGGAGTGAAGTC**GCTGC**CCGTGCTCAAG (5′–3′) TCCTCACTTCAG**CGTCG**GGCACGAGTTC (3′–5′)
G/C duplex	AGGAGTGAAGTC**GCGGC**CCGTGCTCAAG (5′–3′) TCCTCACTTCAG**CGCCG**GGCACGAGTTC (3′–5′)
G/G duplex	AGGAGTGAAGTC**GCGGC**CCGTGCTCAAG (5′–3′) TCCTCACTTCAG**CGGCG**GGCACGAGTTC (3′–5′)
13mer product sequence	AGGAGTGAAGTCG (5′–3′)
15mer product sequence	Phos–CAGCCCGTGCTCAAG (5′–3′)
12mer product sequence	CTTGAGCACGGG (5′–3′)
16mer product sequence	Phos–CTGCGACTTCACTCCT (5′–3′)

Sequence (5'-3')	Melting temperature (T_m) °C
AGGAGTGAAGTC**GCAGC**CCGTGCTCAAG TCCTCACTTCAG**CGTCG**GGCACGAGTTC	76.07
AGGAGTGAAGTC**GCAGC**CCGTGCTCAAG TCCTCACTTCAG**CGACG**GGCACGAGTTC	73

Table 6.2 Oligonucleotides used in this chapter. The highlighted sequences represent the recognition site of TseI. 2=2AP base.

6.2.3 Preparation of 2AP-labelled oligonucleotides

Apart from oligonucleotides containing 2AP, all oligonucleotides were obtained from ATDbio (Southampton, UK). Solid-phase synthesis of 2AP-labelled oligonucleotides was performed using phosphoramidite chemistry[17] on a MerMade DNA/RNA oligonucleotide synthesiser (BioAutomation, USA) in a 5'-DMT-on manner. Purification of the synthesised oligonucleotides was done using standard two-stage DMT-on/DMT-off reverse phase HPLC.[18] The DMT-on full-length products possessed a prolonged retention time during reverse phase HPLC purification and were easily separated from failed sequences. The reaction was detritylated with 40% acetic acid/water and a further DMT-off reverse phase HPLC purification step was applied for higher purity. A NAP-10 column (GE Healthcare) was used for desalting and a Speedvac was used to concentrate the synthesised products. All synthesised oligonucleotides were examined by ESI-FTICR Mass spectrometry to confirm an accurate molecular mass. All the synthesised products were quantified using UV-Vis absorbance at 260 nm with the extinction coefficient determined using Integrated DNA Technologies biophysics online software, which can be found at http://biophysics.idtdna.com/UVSpectrum.html. The emission fluorescence spectra of these three oligonucleotides were also measured using 315 nm excitation wavelength and 10 nm bandpass; the concentration of all samples dissolved in HPLC water was adjusted to 2.5 μM.

6.2.4 Size-exclusion chromatography analysis of TseI

An analytical HPLC gel filtration column calibrated with protein standards was used to determine the molecular weight of TseI in a buffer composed of 20 mM Tris-HCl, 20 mM MES, 10 mM magnesium chloride, 7 mM β-mercaptoethanol, 200 mM sodium chloride, 0.1 mM EDTA, pH 6.5 at room temperature as previously described.[19] TseI was tested at concentrations from 4000 nM to 100 nM.

6.2.5 Measurement of DNA binding to TseI by fluorescent anisotropy

The anisotropy duplex labelled at the 5' end on one strand with HEX was used, Table 6.1. Fluorescence anisotropy measurements were performed using an Edinburgh Instrument FS900 photon counting spectrofluorometer using a T-format measurement and analysed as previously described.[20,21] The excitation wavelength was 535 nm, emission wavelength was 555 nm and bandwidths were 5 nm. 800 μL of 10 nM of the anisotropy duplex were placed in a 10 mm path length quartz cuvette at 25°C. Small amounts of TseI were added from 5 nM to 320 nM to the duplex solution using a microlitre syringe and gently mixed by magnetic stirring. The buffer was 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM calcium acetate, 1 mM DTT, pH 7.9.

6.2.6 Steady-state fluorescence measurements

Five 2AP-labelled oligonucleotides were synthesised, annealed to their complementary strands (Table 6.1) and used for investigating the interaction of TseI with DNA.[22] All duplexes were buffered with 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM DTT, pH 7.9. 200 μL samples of 500 nM duplex plus 1000 nM TseI were incubated for 20 mins at 25°C. Steady-state fluorescence spectra were measured using a FluoroMax (Horiba Jobin Yvon) photon counting spectrofluorometer. Spectra were recorded with a bandpass of 10 nm for both excitation (315 nm) and emission.

6.2.7 Fluorescence-based TseI activity assay

This approach was devised based on previous studies.[23,24] The substrate duplexes consisted of a 5'-HEX-labelled ssDNA annealed to a complementary strand labelled at the 3' end with a BHQ1 group, as shown in Table 6.1. The buffer was 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM DTT, pH 7.9. The duplexes were essentially non-fluorescent in the double-stranded state and highly fluorescent in the single-stranded state achieved by cleavage with TseI at a

temperature greater than the T_m of the cleavage products. The T_m values of duplexes without labelled fluorophore (HEX and BHQ1) are shown in Table 6.1. Based on previous report,[25,26] the author expected that actual T_m values for the duplexes in this experiments were 4-5°C higher than this calculation as the HEX and BHQ1 labels would stabilise the duplexes. The increase of the signal can be easily quantified and initial reaction rates determined. All samples were analysed with an Edinburgh Instrument spectrofluorometer with excitation at 530 nm, emission at 555 nm and 5 nm bandwidths. Each DNA sample (100 μL) was placed in a quartz cuvette with 10 mm path length, put in the fluorometer and allowed to reach the assay temperature of 60°C whereupon a stable fluorescence background signal was achieved. The fluorescence intensity as a function of time was immediately recorded after addition of TseI to a constant monomer concentration of 32.4 nM. Photon counts were converted to the amount of 5'-HEX-labelled oligonucleotide product released using a calibration curve. The substrate concentration versus initial cutting rate was plotted and fitted using the Michaelis-Menten equation[27] ($v=V_{max}[S]/(K_M+[S])$) to determine the kinetic parameters K_M, V_{max} and k_{cat} (also referred as turnover, $k_{cat}= V_{max}/[E]_{total}$) and k_{cat}/K_M (also referred as the specificity constant), where v is the velocity of the reaction, V_{max} is the maximum velocity, [S] is the substrate concentration, and K_M is the substrate concentration at which v is half-maximal ($V_{max}/2$).

6.2.8 Polyacrylamide gel for showing DNA cleavage by TseI

Unlabelled duplexes, as shown in Table 6.1, were used in this experiment at a concentration of 10 μM. 2.5 μL of TseI stock solution (10.8 μM) was added to give a final TseI concentration of 0.26 μM. All samples were incubated at 65°C for ~12 h in the reaction buffer and then run on a 15% polyacrylamide gel in 1×Tris/Borate/EDTA buffer at 150 V, stained with SYBR Gold and viewed under UV light. Single strands corresponding to the reactants and products, as in Table 6.1, were run as markers. The buffer was 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM DTT, pH 7.9.

6.2.9 Denaturing HPLC analysis for showing DNA cleavage by TseI

Unlabelled duplexes, in Table 6.1, were used in this experiment at a concentration of 10 μM. When cleavage was to be tested, 2.5 μL of TseI stock solution (10.8 μM) was added to 100 μL of the duplex solution to give a final TseI concentration of 0.26 μM. All samples were incubated at 65°C for ~12 h. The buffer was 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM DTT, pH 7.9. 20 μL samples were injected onto the HPLC column. The injection solely of TseI in the absence of the duplex gave no HPLC signal above the baseline. A Gilson HPLC equipped with absorption detection at 254 nm was used for analysis of DNA cleavage as reported before.[28] The reverse phase column (C18 Jones Chromatography) was thermostatted at 65°C. A linear acetonitrile gradient from 5% to 65% acetonitrile was generated by mixing 0.1 M acetic acid, 5% acetonitrile and 0.1 M acetic acid, 65% acetonitrile aqueous solution. The pH of the two solvents was adjusted to 6.5 with triethylamine.

6.3 Results and Discussion

6.3.1 Synthesis and characterisation of 2AP-labelled oligonucleotides

The three 2AP-labelled oligonucleotides were successfully synthesized and purified. The preparative HPLC purification profiles are shown in Figure 6.4. The purified samples were further verified by different methods including ESI-FTICR-MS and fluorescence emission spectroscopy as shown in Figure 6.5 and 6.6 respectively.

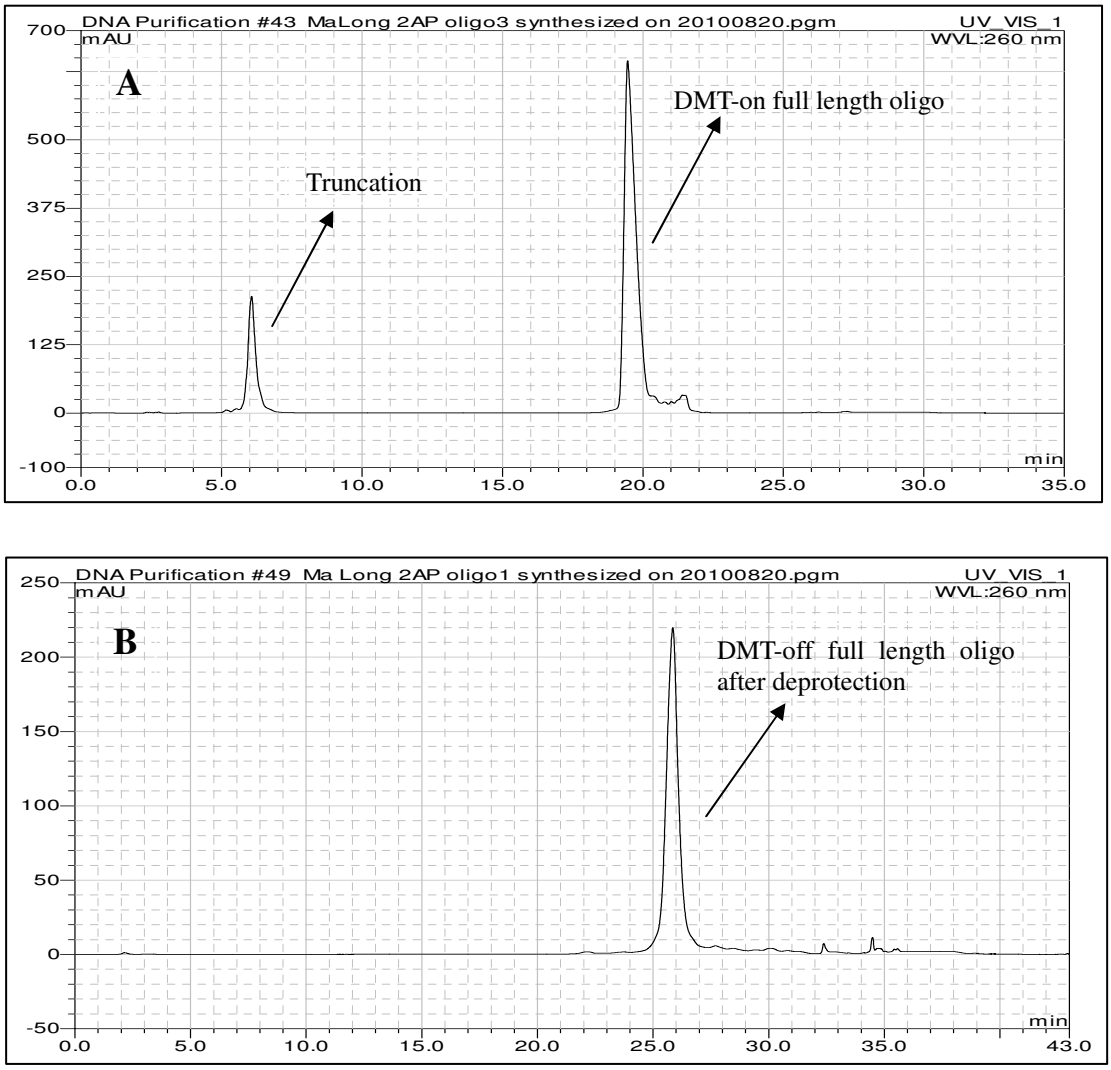


Figure 6.4—Reverse phase HPLC purification profiles of synthesised 2AP-labelled oligonucleotides. (A) DMT-on reverse phase HPLC purification of synthesised 2AP1 (preparative scale). (B) DMT-off reverse phase HPLC purification of synthesised 2AP1 (preparative scale).

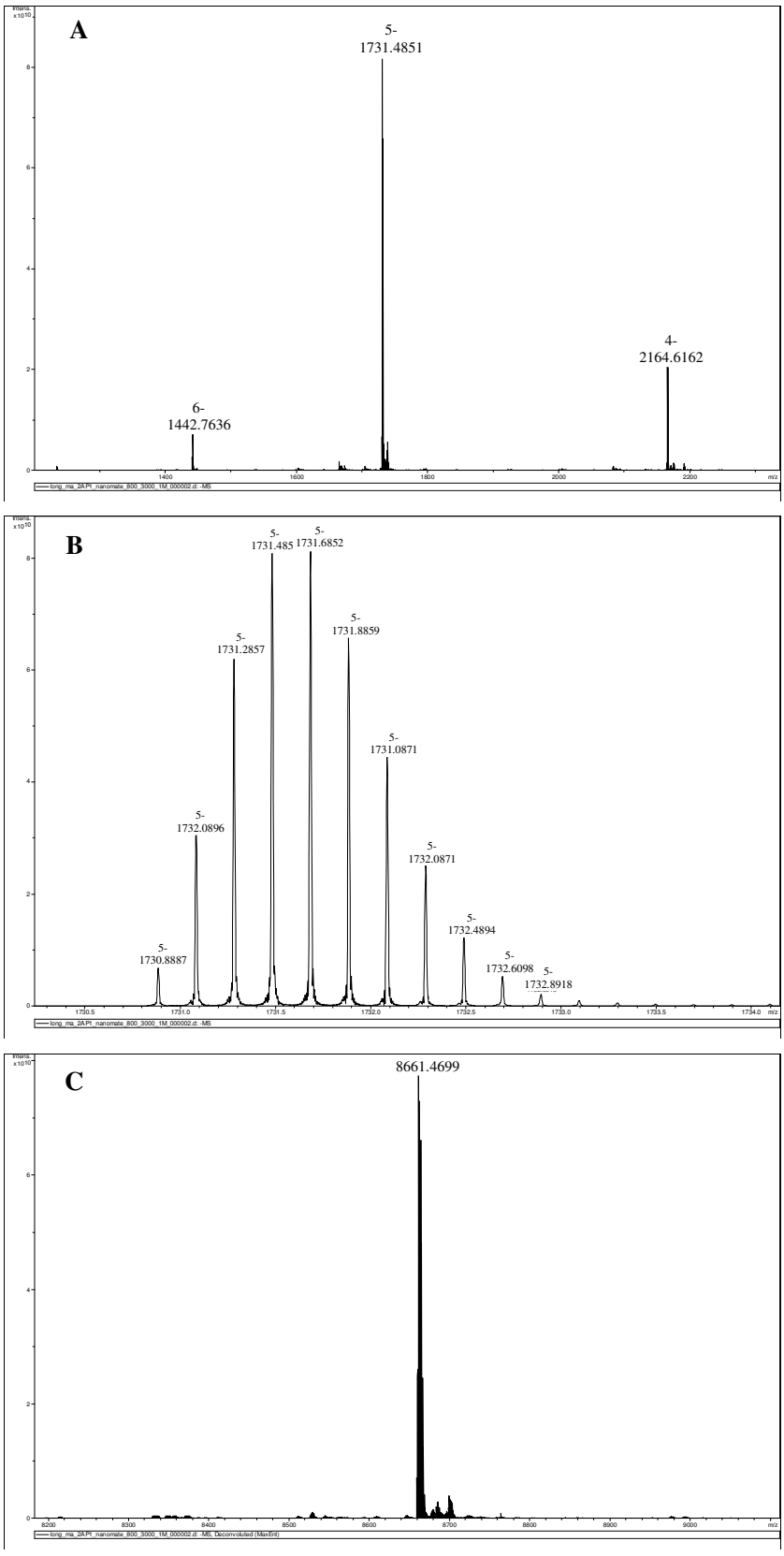


Figure 6.5—ESI-FTICR mass spectra of synthesised 2AP1. (A) Full ESI–MS. (B) 5-charge state ESI–MS. (C) Deconvoluted ESI-MS.

	Sequence	Calculated monoisotopic mass (Da)	M.W. from mass spec (Da)	Composition				Molecular formula
				C	T	A	G	
2AP1	AGGAGTGAAGTCGC2GCCCGTGCTCAAG	8659.47	8659.51	7	4	7	10	$C_{273}H_{314}N_{114}O_{163}P_{27}$
2AP2	CTTGAGCACGGGC2GCGACTTCACTCCT	8251.43	8521.44	10	6	5	7	$C_{270}H_{316}N_{102}O_{167}P_{27}$
2AP3	CTTGAGCACGGGCTGCG2TTCACTCCT	8512.42	8512.42	10	7	4	7	$C_{270}H_{316}N_{102}O_{167}P_{27}$

Table 6.3—Comparison of molecular weights of synthesised ssDNA samples calculated and molecular weights acquired from mass spectrometry. 2=2AP base.

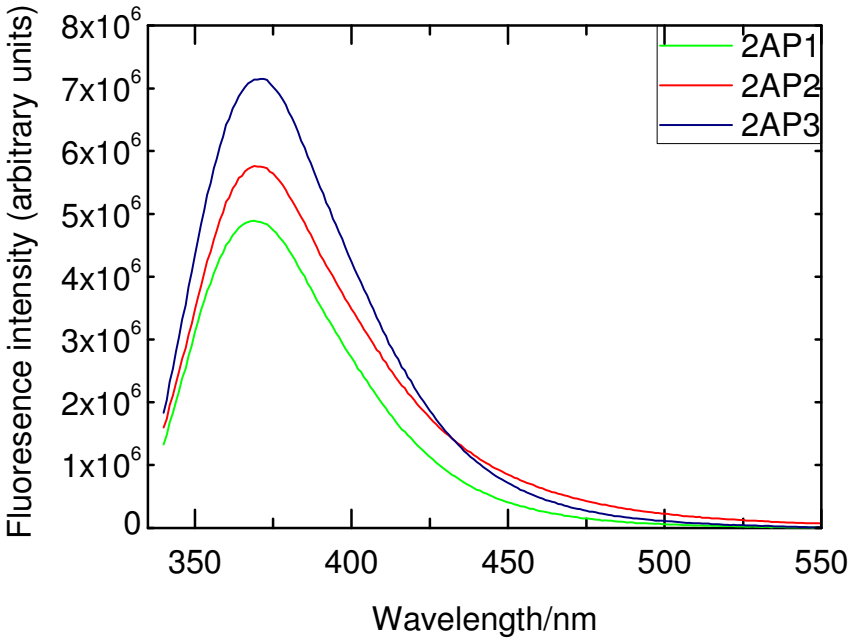


Figure 6.6—Fluorescence spectra of synthesised oligonucleotides. The buffer baseline was subtracted for all spectra.

From the profile of DMT-on HPLC purification, the first HPLC peak with shorter retention corresponds to the truncation product. It is can be seen that a certain amount of truncation product was in the crude synthesised oligonucleotides. The second peak with prolonged retention time corresponds to the full-length product since it is 'DMT-on', which renders it more hydrophobic, resulting in a longer retention time in reverse phase HPLC purification.

The full-length production was detritylated with 40% acetic acid/water and a further 'DMT-off' reverse phase HPLC purification step was done for higher purity. After the subsequent 'DMT-off' purification, a single peak was obtained, indicating a homogeneous fraction of 2AP-labelled DNA. This fraction was collected, desalted concentrated, and quantified. The molecular weight of this product was verified by ESI-FTICR mass spec as shown in Figure 6.5. Figure 6.5A is a full ESI-MS, showing 4-, 5-, 6-charge state species. Figure 6.5B shows the 5-charge species. Figure 6.5C shows the spectrum derived after a deconvolution process. An expected M.W. and experimentally obtained M.W. by MS using the maximum entropy method integrated in the analysis software are listed in Table 6.3. The monoisotopic mass, which is the sum of the masses of the atoms in a molecule using most abundant isotope for each element, is shown in comparison with M.W. from mass spec. They are almost identical within isotope accuracy. Finally, the fluorescent emission spectra of synthesised oligonucleotides were also performed, each spectrum had a characteristic hump peaking at ~370 nm and back to baseline at ~500 nm as reported previously.[29,30] Overall, three 2AP-labelled ssDNA samples were successfully synthesised and purified.

6.3.2 Characterisation of TseI

6.3.2.1 SDS-PAGE for showing the purity of TseI

The purity of TseI enzyme was first checked by SDS-PAGE as shown in Figure 6.7. Protein molecular marker was loaded in the leftmost lane and TseI samples were loaded onto the next three lanes. The amount of TseI loaded onto each line was half the amount as in the previous lane. The SDS-PAGE gel showed a single band in every single lane loaded with different amount of TseI after Coomassie Blue staining corresponding to a molecular weight of ~40 kDa close to that expected from the amino acid sequence of TseI.

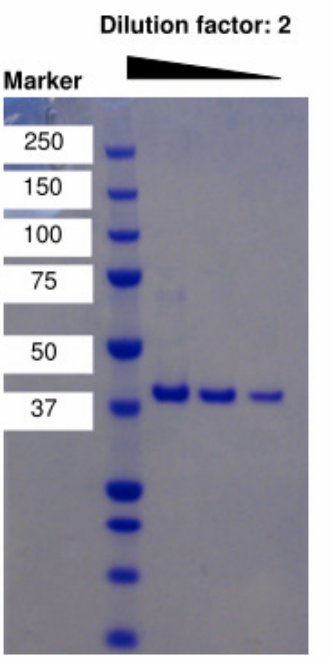


Figure 6.7—4-12% gradient SDS-PAGE gel stained with Coomassie Blue. Lane 1 shows the molecular mass markers (kDa); lanes 2-4 show TseI samples at 5.4 μM, 2.7 μM and 1.35 μM respectively.

6.3.2.2 Size-exclusion chromatography of TseI

Many type II restriction endonucleases are only active in a dimeric form in aqueous solution. In order to confirm the assembly (monomeric or dimeric) of TseI in solution, a calibrated analytical size-exclusion chromatography HPLC was used. The elution profiles of TseI for all concentrations examined ranging from 100 nM to 4000 nM showed a single peak with retention time round about 6.5 min, one example of which (TseI concentration was 1000 nM) is shown in Figure 6.8. The apparent molecular weights can be calculated by the equation derived from the protein standards calibration. Subsequently, the calculated molecular weights against different TseI concentrations were plotted as shown in Figure 6.9, where an average molecular weight of 100±18 kDa was observed. In the concentration range used, this suggests that TseI is a dimer in solution.

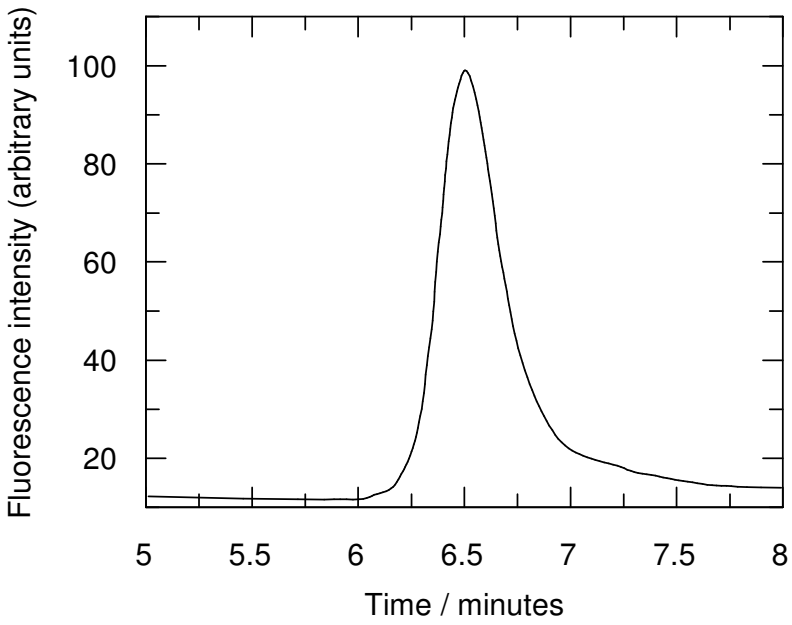


Figure 6.8—Size-exclusion chromatography to investigate the solution molecular mass of TseI. The elution profile (1000 nM shown) shows a single peak at ~6.5 min corresponding to a molecular mass of ~100 kDa.

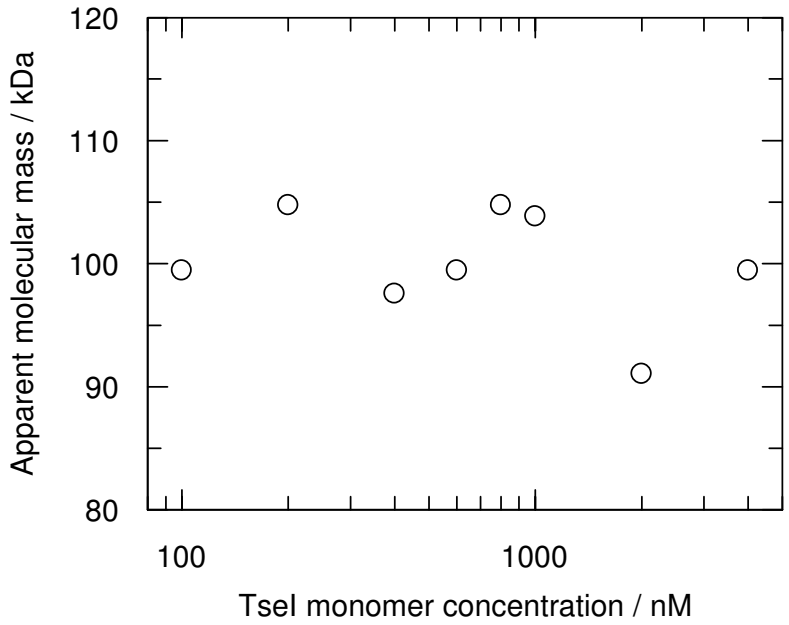


Figure 6.9—Dependence of molecular mass as a function of protein concentration injected onto the column.

6.3.2.3 Fluorescence anisotropy measurements of TseI binding to DNA

The rotational diffusion of the probe on the end of the oligonucleotide is rapid, and the anisotropy of the fluorophore-tagged oligonucleotide is low. The binding of a protein to the DNA significantly decreases the rotational motion of the fluorophore-tagged oligonucleotide, resulting in an increase in the fluorescence emission anisotropy of the tagged DNA as shown in Figure 6.10.

As the intrinsic quantum yield of DNA is too low to be reliably detected, an externally attached fluorescent probe HEX is required. DNA molecules can be chemically modified to label with an extrinsic fluorophore easier than protein molecules. HEX is a widely used extrinsic fluoroscent dye and the fluorescence signal of it can be monitored to study the "tumbling" of the DNA molecule in solution.[31,32] The HEX probe was chemically attached to the ssDNA *via* a six-carbon linker and then the ssDNA was annealed with its complementary strand. This probe is neither too flexible nor too rigid compared with the DNA molecule therefore the motion of the conjugated HEX dye is able to mirror the motion of the DNA molecule of interest.[33] A slight degree of freedom was detected; however it is not enough to preclude the use of the labelled duplex for the study of protein-DNA interactions. Moreover, HEX has a satisfactory quantum yield, enabling it to be used at very low concentrations, which is crucial advantage because tight protein-DNA binding needs to be analysed at very low DNA and protein concentrations.[34]

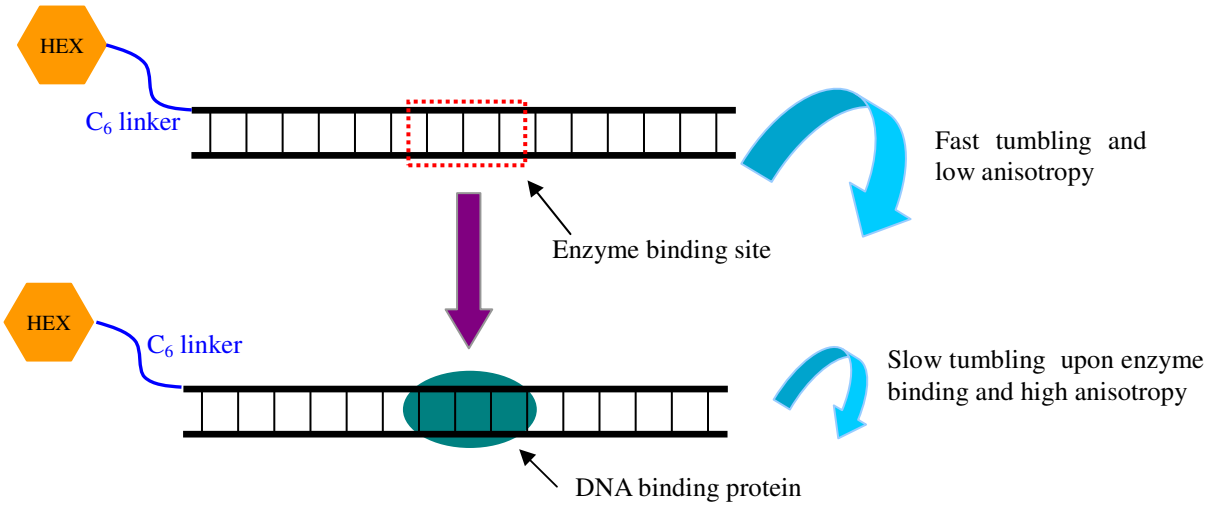


Figure 6.10—Illustration of principle for monitoring binding by fluorescence anisotropy (Adapted from Lundblad *et al.*[34]). DNA is "tumbling" in solution in a faster rate when free from bound, and showing a lower anisotropy of the HEX extrinsic fluorophore. The binding of protein molecule to the DNA decreases the mobility of the DNA, and hence less "tumbling" is presented, giving rise to an increase in the anisotropy of the labelled HEX.

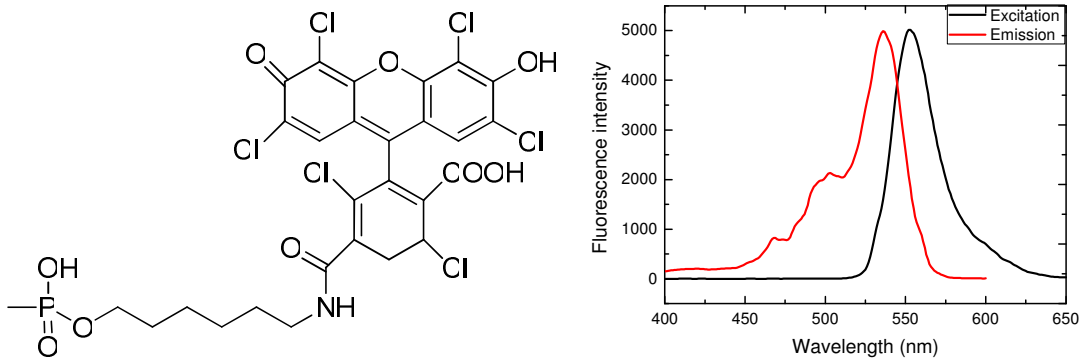


Figure 6.11—Chemical structure and fluorescent spectra of HEX, the fluorescent dye used for TseI anisotropy measurements. The left panel shows the chemical structure of HEX; the right panel shows the characteristic excitation (red) and emission (black) spectra of HEX-labelled ssDNA.

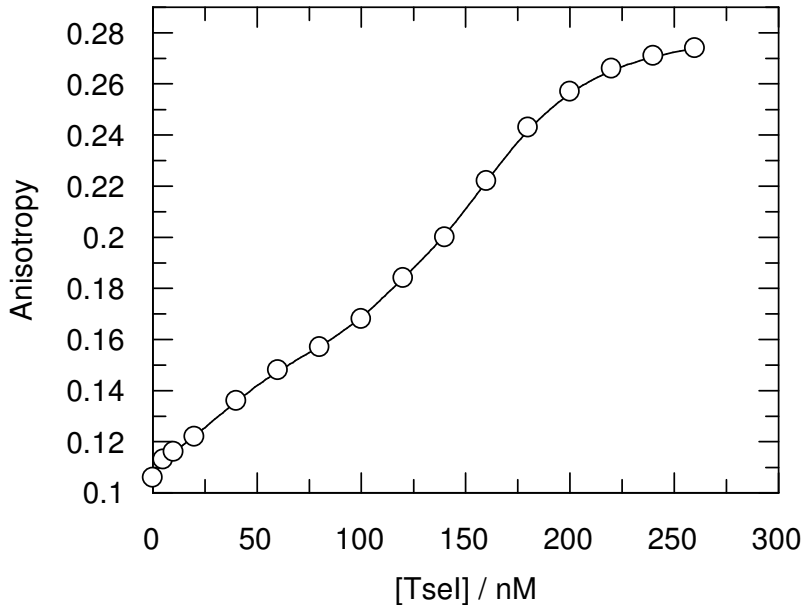
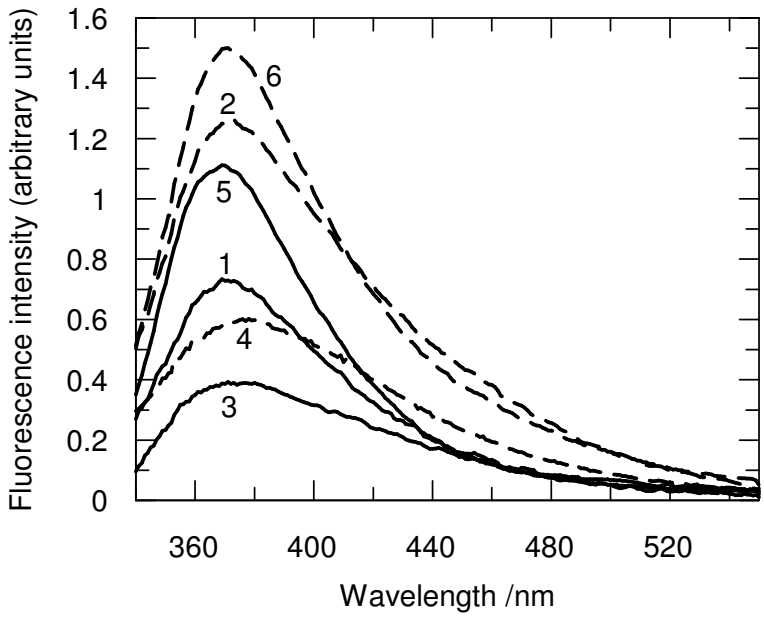


Figure 6.12—Fluorescence anisotropy increase for 10 nM HEX-labelled 28 bp DNA duplex as a function of TseI monomer concentration (data obtained in single experiment). In the absence of enzyme, the anisotropy had a value of 0.106 and this increased to ~0.275 when all DNA was bound.

Binding to a HEX-labelled duplex containing the target sequence for TseI was examined using the increase in fluorescence anisotropy of the HEX label caused by the slowing of molecular rotation when the mass of the duplex was increased by protein binding as shown in Figure 6.12. Binding did take place when adding TseI into the solution and complete binding was observed at a TseI monomer concentration of 300 nM and 50% binding at ~120 nM. However, the binding appeared to be biphasic with a higher affinity binding site, with an estimated dissociation constant around ~40 nM. The initial and final values of the anisotropy were similar to those previously observed for HEX-labelled DNA binding to M.EcoKI, an enzyme of molecular weight 170 kDa.[35]

6.3.2.4 Steady-state of fluorescence intensity measurements

Several restriction enzymes recognising DNA target sequences similar to that recognised by TseI have been shown by crystallography and 2AP fluorescence studies to flip out the central base pair in the target sequence and to collapse the DNA duplex thus converting a 5 bp recognition sequence into a four base pair recognition sequence.[10,11,30,36] Steady-state measurement of 2AP has been becoming a readily used method for monitoring nucleotide unstacking.[13,37-39] 2AP is highly quenched within duplex DNA mostly due to the base stacking interactions. The quantum yield of 2AP is susceptible to the outside environment. Strong base unstacking effects like base flipping can result in dramatically increased fluorescence intensity. The intensity increases in 2AP emission due to enzyme binding and base flipping ranges from 5-fold to 150-fold for these kinds of enzymes.[13,30]



igure 6.13—Steady-state fluorescence emission spectra of 2AP-labelled DNA in the absence/presence of TseI enzyme. The solid lines represent the fluorescence intensity of DNA in the absence of TseI for duplexes 2AP/T duplex 1 (spectrum 1), A/2AP duplex 2 (spectrum 3) and T/2AP duplex 3 (spectrum 5). The dotted lines represent the fluorescence intensity of the same duplexes in the presence of excess TseI (spectra 2, 4 and 6 respectively).

No structure is available for TseI thus the fluorescence-based approach was used for testing the possibility of the base flipping mechanism adopted by TseI. This was done by replacing the central bases with 2AP either paired with T (2AP:T) or as a mismatch with A (2AP:A). As a control, 2AP was also placed outside of the target sequence and paired with T. The fluorescence of the duplexes containing 2AP showed a typical 2AP emission spectrum when excited at 315 nm with emission maxima at ~370 nm, as shown in Figure 6.13. The addition of excess TseI to ensure near complete binding of the duplex only caused a small increase (less than 50%) in the emission intensity without obvious changes in the emission maximum wavelength. Most importantly, the same kinds of change in intensity were observed for all locations and all base pairings of the 2AP probe. Thus it appears that TseI does not use the base flipping mechanism in its recognition of its DNA target and the intensity changes observed must be due to a minor base disturbance upon enzyme binding.

6.3.2.5 A continuous fluorescence-based assay

The endonuclease activity of TseI on short DNA duplexes containing the target sequence or the target sequence having a mismatched base pairing in the centre of the recognition sequence was investigated using a continuous fluorescence assay. The assay uses the difference in thermal stability of the 28 bp substrate and the shorter products as initially proposed by Waters *et al.*[40] who used the increase in absorption due to the melting of the shorter cleaved products. This was subsequently adapted for fluorescence measurements for measuring DNA cleavage by Thorson *et al.*[23] who used the melting of the shorter products to remove a fluorescence quencher on the 3' end of one strand from contact with a fluorescence reporter on the 5' end of the other strand. Some others, as Tan *et al.*[41] and Pingoud *et al*.[42] adopted similar approaches to monitor enzymatic cleavage of DNA. The T_m of the fully matched and A:A mismatched duplexes were calculated to be 76°C and 73°C respectively. As long as the assay temperature lies between the T_m of the substrate and the products, the products melt upon cleavage by the enzyme and the fluorescence of the fluorophore is greatly increased. It was indicated by Tyagi *et al*.[25] that the T_m of blunt-ended

dsDNA containing fluorophore-quencher (HEX-BHQ1) can be enhanced by 4-5°C, which can ensure the duplexes in the study were fully double-stranded at the temperature (60°C) when performing experiments. This assay worked very well for TseI, as shown in Figure 6.14B, a substantial increase in fluorescence from a low background level as a function of time after addition of the enzyme. The initial rate of fluorescence increase was determined as a function of substrate concentration for both of the duplexes, one containing the target sequence and the other one containing an A:A mismatch in the middle of the sequence. Both duplexes were cleaved by the enzyme and in fact the mismatched duplex was cleaved faster as shown in the Michaelis-Menten analysis, Figure 6.14C. The maximum velocity for cleavage was 3.3±0.2 nM s^{-1} and 6.0±0.8 nM s^{-1} for the matched duplex and the mismatched duplex respectively. k_{cat} values were 0.20±0.01 s^{-1} and 0.37±0.05 s^{-1} for the normal duplex and the mismatched duplex respectively assuming that 100% of the enzyme molecules were active as dimers. The Michaelis-Menten constant, K_M, was 49.6±11.5 nM and 80.7±30.6 nM for the matched and mismatched duplexes respectively. The values of k_{cat}/K_M were $4.0\pm0.95\times10^{-3}$ nM^{-1} s^{-1} and $4.6\pm3.4\times10^{-3}$ nM^{-1} s^{-1}, indicating the enzyme has no preference for one substrate over the other.

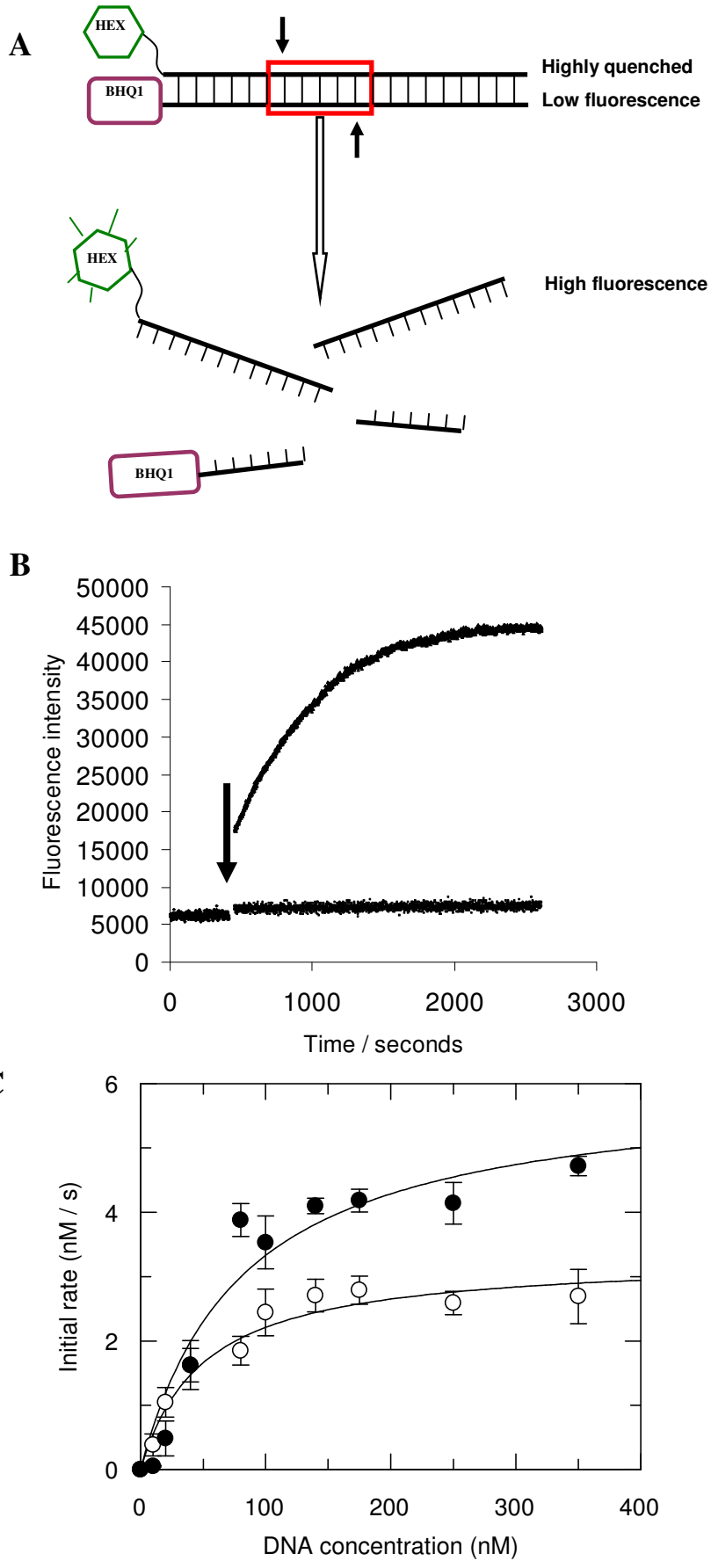

A
HEX
BHQ1
Highly quenched
Low fluorescence
HEX
High fluorescence
BHQ1
B
Fluorescence intensity
50000
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
0
1000
2000
3000
Time / seconds
C
Initial rate (nM / s)
6
4
2
0
0
100
200
300
400
DNA concentration (nM)

Figure 6.14—Fluorescence-based TseI activity assay. (A) 5'-HEX labelled top strand was annealed with 3'-BHQ1 labelled strand (ratio of 5'-HEX labelled to 3'-BHQ1 labelled strand=1:1.2) and became highly quenched. Adding TseI at elevated temperature (60℃) caused cleavage of the duplex resulting in the separation of the fluorophore-quencher pair and therefore appearance of fluorescence. (B) Fluorescence intensity as a function of time. Initially a low signal was observed. At the time indicated by the arrow an aliquot of either TseI or buffer was added. TseI caused the rapid increase in signal but the buffer addition caused no further change in signal. (C) Michaelis-Menten plots of TseI duplex cutting reaction for both matched (○) and A:A mismatched DNA substrate (●). Error bars are standard deviations for experiments performed in triplicate.

6.3.3 Cleavage of TseI on matched and mismatched DNA sequences

As shown by using a continuous fluorescence-based assay, TseI appeared to be capable of cleaving A:A mismatch in the central position of its recognition sequence. In this section, the author designed a series of mismatches in the central base of its 5 bp recognition site, trying to further verify this specific property of TseI using PAGE and denaturing HPLC.

6.3.3.1 Gel electrophoresis results for visualising DNA cleavage by TseI

Figure 6.15 shows that 28 bp duplexes containing mismatches (A:A, T:T) were all cleaved into shorter duplexes as effectively as the normal cognate sequence but that those duplexes containing G:G or G:C at the central position were not cleaved as expected since they lacked the target sequence.

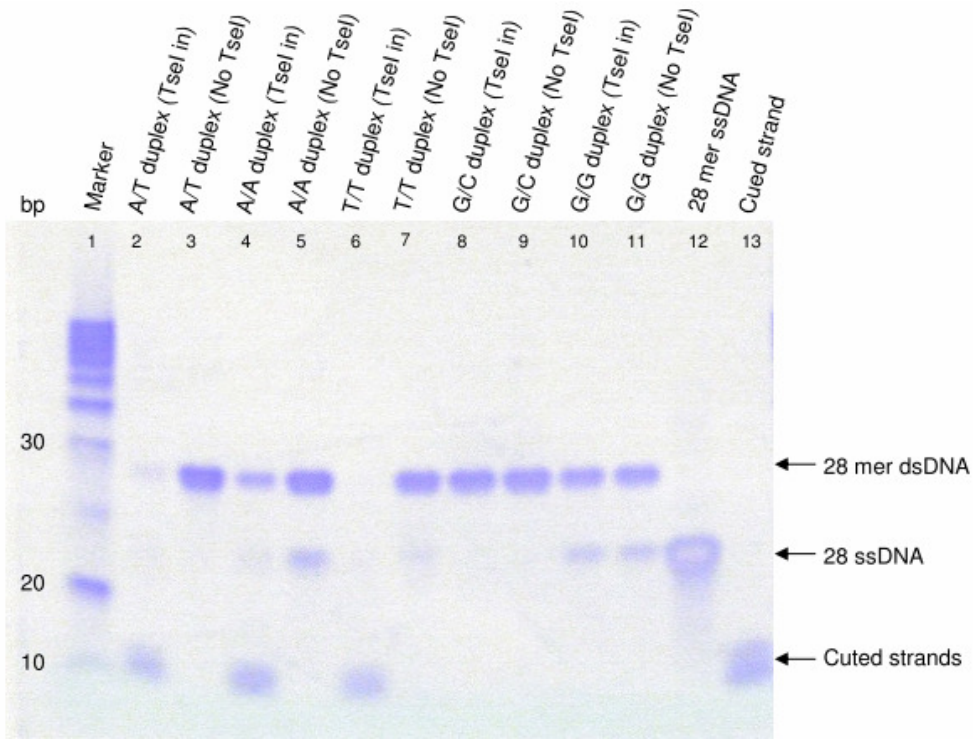
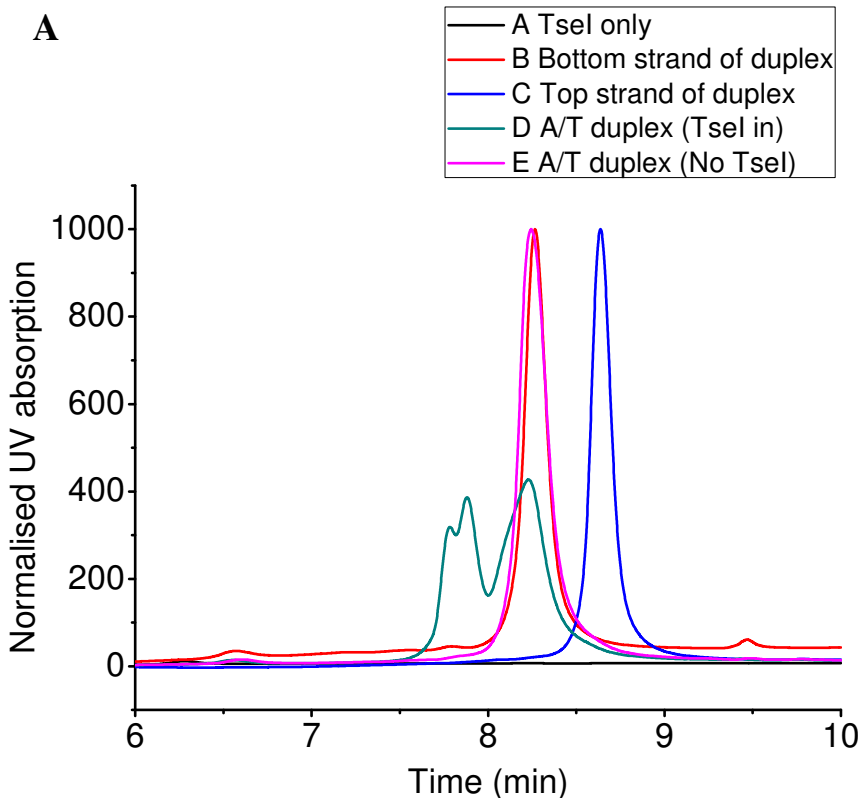


Figure 6.15—Polyacrylamide gel for analysing matched and mismatched DNA cleavage by TseI. Lane 1 was the molecular mass marker; lane 3, 5, 7 were the duplexes in the absence of TseI (negative control); lane 2, 4, 6 were the duplex in the presence of TseI; lane 12 was a 28 mer ssDNA (top strand of A/T duplex); lane 13 was the mixture of four TseI A/T duplex cutting product strands. Lanes 2, 4, 6 clearly show DNA cleavage by TseI, yielding four 12-16 mer ssDNA.

6.3.3.2 Denaturing HPLC experiments for analysing DNA cleavage by TseI

The ability of the enzyme to cleave a target sequence containing a mismatch was unexpected and would allow a new use for TseI in investigating A:A and T:T mismatches formed by the formation of hairpins in repetitive DNA sequences such $(CAG)_n$ found in many genetic diseases. Thus mismatched duplexes cleavage was further explored.

Rather than performing full enzyme activity studies on each possible mismatch using the fluorescence assay (due to the expense of the two labels), gel assays, as shown in Figure 6.15, and HPLC assays as shown in Figure 6.16-Figure 6.18, for cleavage were performed on duplexes containing A:T, A:A, T:T, G:G, G:C, A:2AP *etc*. in the central position of the target sequence. Figure 6.16 and 6.17 show that 28 bp duplexes containing mismatches of A or T were all cleaved into shorter duplexes as effectively as the normal cognate sequence but that those duplexes containing G or C at the central position were not cleaved as expected since they failed to contain the target sequence.



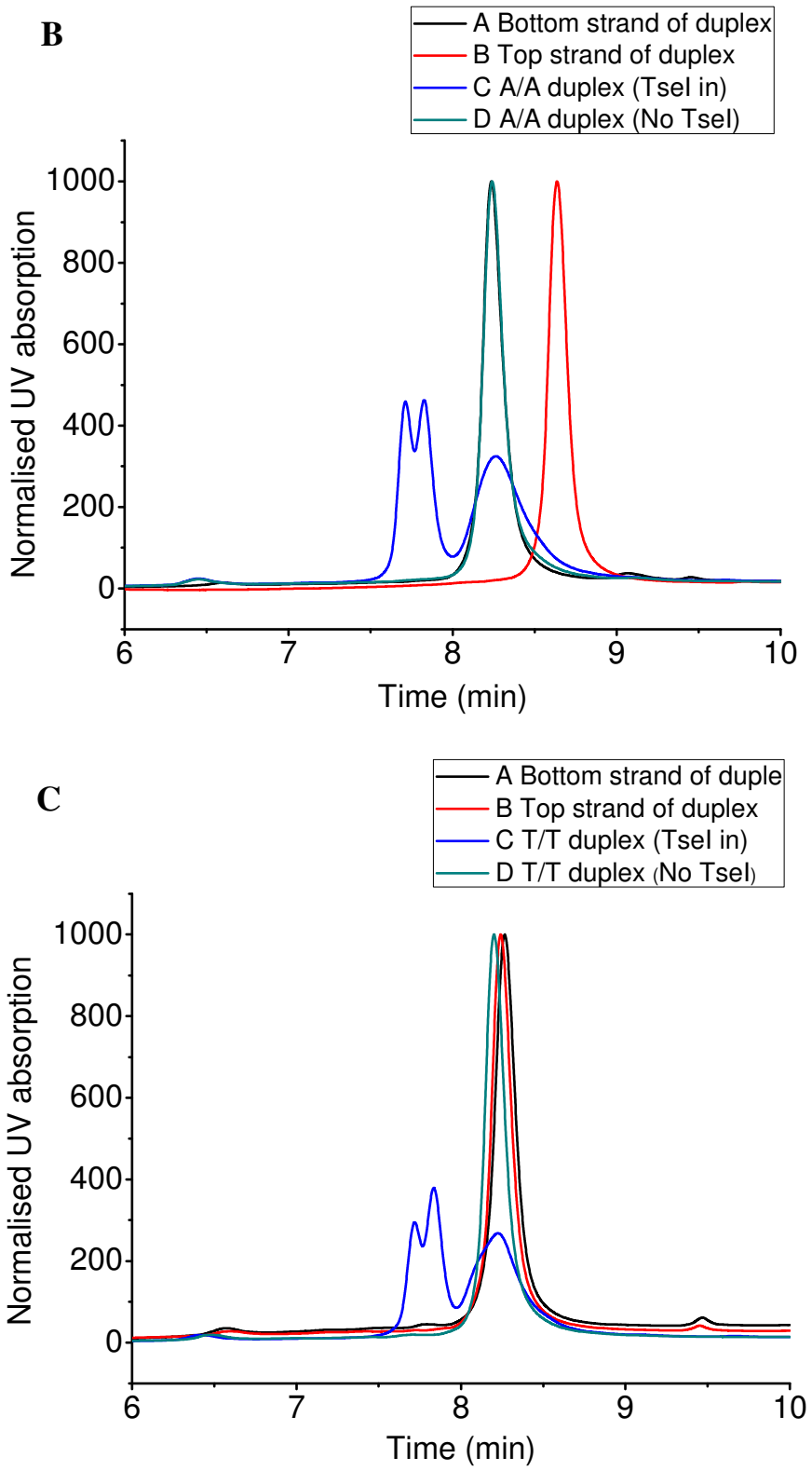


Figure 6.16—HPLC analysis of the action of TseI on DNA duplexes. DNA duplex with A:T base pair (A), A:A base pair (B), and T:T base pair (C) in central position of TseI recognition site.

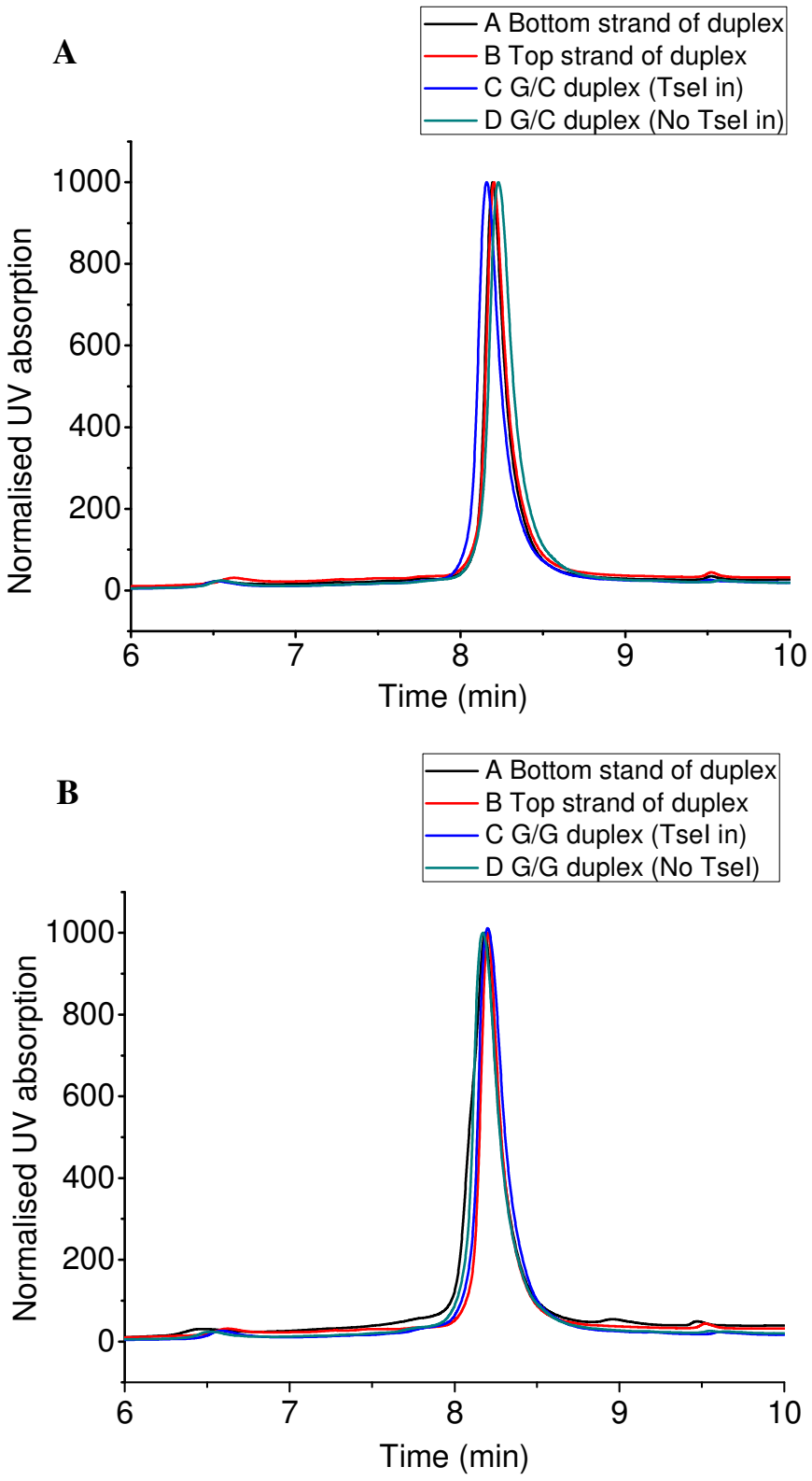


Figure 6.17—HPLC analysis of the action of TseI on DNA duplexes. DNA duplex with G:C base pair (A) and G:G base pair (B) in central position of TseI recognition site. For (A) and (B), they contain elution profiles of bottom and top strands, elution profiles of duplexes with and without TseI respectively.

HPLC analysis of duplex substrates and products using a reverse phase chromatography column run at elevated temperature to denature the DNA was performed.[43] Results show that the substrate duplexes did not denature on the column and eluted as a single peak for all substrates investigated. The shorter product strands (resulted from TseI cleavage) eluted as poorly resolved peaks at an elution time substantially lower than that of the uncleaved DNA strands as shown in Figure 6.16 A-C and thus cleaved DNA could be clearly distinguished from uncleaved DNA. Figure 6.16 clearly demonstrated that duplexes containing the normal target sequence or A:A or T:T mismatches in the central position of TseI recognition sequence were cleaved, However, as shown in Figure 6.17, the duplexes lacking the target site were not cleaved if G:G or C:C mismatches were present at the central base pair of the target sequence. Moreover, the retention times of cleaved DNA chromatographic peaks were comparable to those of the product ssDNA (shown in Figure 6.18). The retention times of the poorly resolved peaks in Figure 6.16A-C corresponded four standard ssDNA samples (12, 13, 15, 16 mer) as shown in Figure 6.18very well, indicating cleavage.

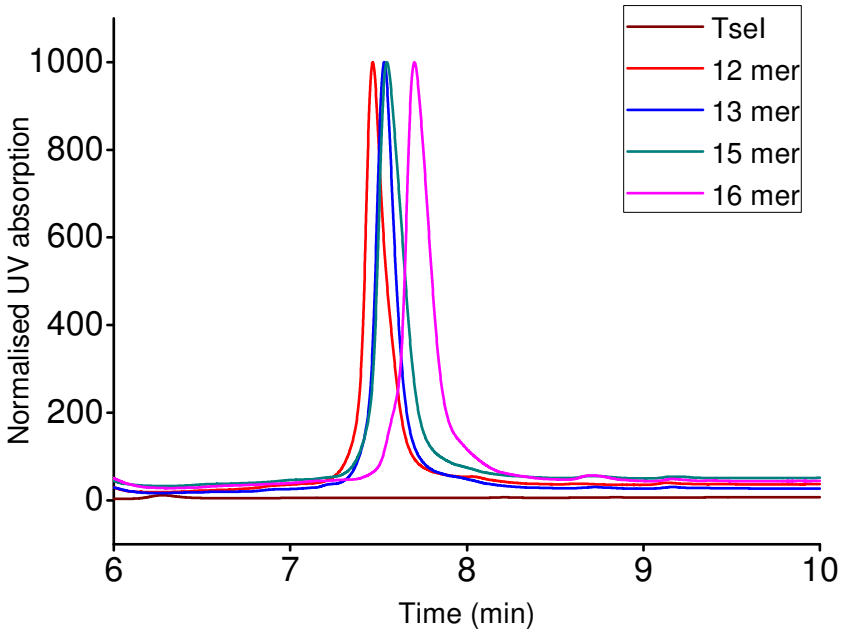


Figure 6.18—HPLC analysis of the action of TseI on DNA duplexes. Elution profiles of ssDNA (12, 13, 15 and 16 mer as shown in Table 6.1) and TseI enzyme only.

Based on previous publications, there are a range of type II endonucleases capable of cleaving recognition sites containing mismatched base pair. According to Palladino *et al.*[44] two non-radioactive assays were used to evaluate cleavage of heteroduplex DNA containing mismatches. These assays demonstrated that some of restriction enzymes under investigation were capable of site-specific double-stranded cleavage with mismatched heteroduplex DNA substrates including BamHI, EcoRV, HindIII, PstI, SacI, SalI. However, certain mismatched substrates did effectively abolish cleavage to undetectable levels. Jiricny *et al.*[45] found that restriction endonuclease HindII and TaqI can efficiently cleave synthetic hexadecanucleotide duplexes which contained either an A:C or a G:T mismatch within their respective restriction sites. Also, HindII and TaqI were found capable of cleaving DNA sequences with one base substituted with guanosine analogues inosine and 7-deazainosine, or with adenosine analogues tubercidin in their recognition sites.[46] Petranović *et al.*[47] found that SmaI and XmaI cleaved both T:G and C:A mismatch-containing sequences; AvaI, HpaII, MspI, NciI, NspIII were able to cleave only the T:G containing sequence; while BstNI was able to cleave only the C:A containing sequence. All these results show that cleaving mismatches is not unique for TseI, and the use of this characteristic feature can be possibily expanded in certain aspect such as DNA mismatch hunting, DNA mismatch repair *etc*.

The author experimentally demonstrated that TseI is capable of cleavage of A:A and T:T mismatches in the central position of its recognition site with no preference for matched sequence over the central mismatched sequence at elevated temperature. Also according to Duzdevich *et al.*[9] TseI preferentially cleaves normal duplex DNA at ambient temperature, and digested the mismatches caused by super-long CAG repeats only at higher temperatures. This may be because the unusual structure of DNA containing CAG expansions sterically hinders the access of the restriction enzyme to DNA. If so, by judicious choice of incubation conditions, it should be possible to use the enzyme to discriminate between the two forms of DNA. For this reason, the author suggests that TseI should be a useful tool for exploring the role of DNA structure in trinucleotide repeat disorder such as HD (CAG repeats) and myotonic dystrophy type 1 (CTG repeats).

6.4 References

1. A. Pingoud, M. Fuxreiter, V. Pingoud and W. Wende, *Cell. Mol. Life Sci.*, 2005, 62, 685-707.
2. A. Pingoud and A. Jeltsch, *Nucleic Acids Res.*, 2001, 29, 3705-3727.
3. T. A. Bickle and D. H. Krüger, *Microbiol. Rev.*. 1993, 57, 434-450.
4. H.-C. Guo, *Recent Res. Dev. Macromol.*, 2003, 7, 225-245.
5. R. J. Roberts, T. Vincze, J. Posfai and D. Macelis, *Nucleic Acids Res.*, 2005, 33, D230-D232.
6. R. J. Roberts, T. Vincze, J. Posfai and D. Macelis, *Nucleic Acids Res.*, 2010, 38, D234-D236.
7. R. J. Roberts, M. Belfort, T. Bestor, A. S. Bhagwat, T. A. Bickle, J. Bitinaite, R. M. Blumenthal, S. K. Degtyarev, D. T. F. Dryden, K. Dybvig, K. Firman, E. S. Gromova, R. I. Gumport, S. E. Halford, S. Hattman, J. Heitman, D. P. Hornby, A. Janulaitis, A. Jeltsch, J. Josephsen, A. Kiss, T. R. Klaenhammer, I. Kobayashi, H. Kong, D. H. Krüger, S. Lacks, M. G. Marinus, M. Miyahara, R. D. Morgan, N. E. Murray, V. Nagaraja, A. Piekarowicz, A. Pingoud, E. Raleigh, D. N. Rao, N. Reich, V. E. Repin, E. U. Selker, P.-C. Shaw, D. C. Stein, B. L. Stoddard, W. Szybalski, T. A. Trautner, J. L. Van Etten, J. M. B. Vitor, G. G. Wilson and S. y. Xu, *Nucleic Acids Res.*, 2003, 31, 1805-1812.
8. R. R. Sinden, V. N. Potaman, E. A. Oussatcheva, C. E. Pearson, Y. L. Lyubchenko and L. S. Shlyakhtenko, *J. Biosci.* , 2002, 27, 53-65.
9. D. Duzdevich, J. Li, J. Whang, H. Takahashi, K. Takeyasu, D. T. F. Dryden, A. J. Morton and J. M. Edwardson, *PLoS ONE*, 2011, 6, e17119.
10. M. Bochtler, R. H. Szczepanowski, G. Tamulaitis, S. Grazulis, H. Czapinska, E. Manakova and V. Siksnys, *EMBO J.*, 2006, 25, 2219-2229.
11. R. K. Neely, G. Tamulaitis, K. Chen, M. Kubala, V. Siksnys and A. C. Jones, *Nucleic Acids Res.*, 2009, 37, 6859-6870.
12. G. Tamulaitis, M. Zaremba, R. H. Szczepanowski, M. Bochtler and V. Siksnys, *Nucleic Acids Res.*, 2008, 36, 6101-6108.
13. M. A. Carpenter and A. S. Bhagwat, *Nucleic Acids Res.*, 2008, 36, 5417-5425.
14. E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel and A. Bairoch,

Nucleic Acids Res., 2003, 31, 3784-3788.

15. W. A. Kibbe, *Nucleic Acids Res.*, 2007, 35, W43-W46.
16. R. Owczarzy, A. V. Tataurov, Y. Wu, J. A. Manthey, K. A. McQuisten, H. G. Almabrazi, K. F. Pedersen, Y. Lin, J. Garretson, N. O. McEntaggart, C. A. Sailor, R. B. Dawson and A. S. Peek, *Nucleic Acids Res.*, 2008, 36, W163-W169.
17. S. L. Beaucage and M. H. Caruthers, *Tetrahedron Lett.*, 1981, 22, 1859-1862.
18. K. Itakura, J. J. Rossi and R. B. Wallace, *Annu. Rev. Biochem.*, 1984, 53, 323-356.
19. D. T. F. Dryden, L. P. Cooper, P. H. Thorpe and O. Byron, *Biochemistry*, 1997, 36, 1065-1076.
20. L. M. Powell, D. T. F. Dryden, D. F. Willcock, R. H. Pain and N. E. Murray, *J. Mol. Biol.*, 1993, 234, 60-71.
21. M. O'Neill, D. T. F. Dryden and N. E. Murray, *EMBO J.*, 1998, 17, 7118-7127.
22. L. M. Wilhelmsson, *Q. Rev. Biophys.*, 2010, 43, 159-183.
23. J. B. Biggins, J. R. Prudent, D. J. Marshall, M. Ruppen and J. S. Thorson, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, 97, 13537-13542.
24. S. S. Ghosh, P. S. Eis, K. Blumeyer, K. Fearon and D. P. Millar, *Nucleic Acids Res.*, 1994, 22, 3155-3159.
25. S. A. E. Marras, F. R. Kramer and S. Tyagi, *Nucleic Acids Res.*, 2002, 30, e122.
26. B. G. Moreira, Y. You, M. A. Behlke and R. Owczarzy, *Biochem. Biophys. Res. Commun.*, 2005, 327, 473-484.
27. K. A. Johnson and R. S. Goody, *Biochemistry*, 2011, 50, 8264-8269.
28. T. J. Su, B. A. Connolly, C. Darlington, R. Mallin and D. T. F. Dryden, *Nucleic Acids Res.*, 2004, 32, 2223-2230.
29. E. Y. M. Bonnist and A. C. Jones, *ChemPhysChem*, 2008, 9, 1121-1129.
30. G. Tamulaitis, M. Zaremba, R. H. Szczepanowski, M. Bochtler and V. Siksnys, *Nucleic Acids Res.*, 2007, 35, 4792-4799.
31. M. U. Kumke, H. G. Löhmannsröben and T. Roch, *J. Fluoresc.*, 1995, 5, 139-152.
32. F. W. Sevenich, J. r. Langowski, K. Rippe and V. Weiss, *Nucleic Acids Res.*,

1998, 26, 1373-1381.

33. L. M. Powell, B. A. Connolly and D. T. F. Dryden, *J. Mol. Biol.*, 1998, 283, 947-961.
34. J. R. Lundblad, M. Laurance and R. H. Goodman, *Mol. Endocrinol.*, 1996, 10, 607-612.
35. L. M. Powell, D. T. F. Dryden and N. E. Murray, *J. Mol. Biol.*, 1998, 283, 963-976.
36. R. H. Szczepanowski, M. A. Carpenter, H. Czapinska, M. Zaremba, G. Tamulaitis, V. Siksnys, A. S. Bhagwat and M. Bochtler, *Nucleic Acids Res.*, 2008, 36, 6109-6117.
37. B. Holz, E. Weinhold, S. Klimasauskas and S. Serva, *Nucleic Acids Res.*, 1998, 26, 1076-1083.
38. B. W. Allan and N. O. Reich, *Biochemistry*, 1996, 35, 14757-14762.
39. Y. V. R. Reddy and D. N. Rao, *J. Mol. Biol.*, 2000, 298, 597-610.
40. T. R. Waters and B. A. Connolly, *Anal. Biochem.*, 1992, 204, 204-209.
41. J. J. Li, R. Geyer and W. Tan, *Nucleic Acids Res.*, 2000, 28, e52.
42. K. Eisenschmidt, T. Lanio, A. Jeltsch and A. Pingoud, *J. Biotechnol.*, 2002, 96, 185-191.
43. W. Xiao and P. J. Oefner, *Hum. Mutat.*, 2001, 17, 439-474.
44. M. T. Langhans and M. J. Palladino, *Curr. Issues Mol. Biol.*, 2009, 11, 1-12.
45. J. Jiricny and D. Martin, *Nucleic Acids Res.*, 1986, 14, 1943-1949.
46. J. Jiricny, S. G. Wood, D. Martin and A. Ubasawa, *Nucleic Acids Res.*, 1986, 14, 6579-6590.
47. M. Petranović, D. Petranović, C. Dohet, P. Brooks and M. Radman, *Nucleic Acids Res.*, 1990, 18, 2159-2162.

Chapter 7---General Remarks and Future Work

All the work presented in the thesis can be divided into two avenues: a study focused on enzymology using biochemical and biophysical methods and secondly, the study of DNA transient dynamics either existing as a nucleic acid structure *per se* or induced by nucleic acid enzyme binding *in vitro* using time-resolved fluorescence measurements.

As shown in chapter 3, time-resolved fluorescence demonstrated its competence in probing transient unzipping in DNA structure. Subsequently by using this method, a systematic study was performed to look at the dynamics of bases in different positions (either in or near the branch point or in the helical arms) of a perfectly base-paired DNA 3WJ. It was found that the bases adjacent to the branch point of 3WJ are unzipped, despite the full Watson-Crick complementarity of the molecule. The unzipping could allow a nanoscale cavity to form at the junction centre. It is anticipated that these results will guide the development of new DNA-based supramolecular receptors and nanosystems.

By utilising the rationalisation developed in chapter 3, the author used the time-resolved fluorescence of 2AP to examine the double-nucleotide unzipping mechanism adopted by human flap endonuclease for achieving accurate DNA incision in chapter 4. It was found through comparison amongst the response of the 2AP base in the +1 position and -1 position (+1 and -1 is the position with respect to scissile phosphodiester) in complex with WT hFEN1, Y40A, R100A, K93A, that WT hFEN1 does unzip the double-base prior to DNA cleavage. Moreover, Y40 stacks with the +1 and -1 base in the WT hFEN1:substrate DNA complex. While in Y40A, R100A and K93A, the unzipping effect can't be obviously observed. It was concluded that double-nucleotide unzipping occurs as a prerequisite for catalysis and achieving the specificity.

In chapter 5, the heteromultimeric type III restriction enzyme EcoP15I was studied. It was found that it adopts a Res_1Mod_2 stoichiometry, which is in conflict to the Res_2Mod_2 stoichiometry reported previously. The base flipping mechanism of EcoP15I was tested, and it was found that it flips the second adenine base (also is

the methyl transfer target) in its recognition site 5'-CAGCAG-3', rather than the first one. Lastly, the DNA cleavage property of EcoP15I on a substrate containing two inversely oriented recognition sites was investigated, in order to find the requirement for EcoP15I to achieve efficient DNA cleavage. It requires at least one fully matched 5'-CAGCAG-3' sequence for two inversely located adjacent recognition sites.

In chapter 6, the type II restriction enzyme TseI was biochemically and biophysically characterised. Also, the mismatched DNA cleavage feature (capable of cleaving the duplexes having A:A or T:T mismatch in the central position of the 5'-GCWGC-3' recognition site, W=A, T) was validated by a fluorescence-based enzyme assay, analytical denaturing HPLC and gel assays. This feature could be used as a useful tool for exploring DNA structure containing super-long CAG or CTG repeats found in some trinucleotide repeat disorders. The base flipping was also tested for TseI, but it can be concluded that it doesn't adopt a base flipping mechanism when performing the enzymatic reaction.

All the work done in this thesis falls in the DNA-protein interactions domain. DNA-protein complexes can be studied at three distinct levels: at DNA level, protein level and the complex level. Specific interactions between proteins and DNA are fundamental to many biological processes, providing an extremely crucial role for life maintenance, evolution and function performing. Studying those interactions would be of help to people in understanding things taking place in biology during growth, development, differentiation, evolution *etc.* Using ultrafast time-resolved spectroscopy, the temporal resolution can be used to capture the biologically relevant conformational dynamics of substrate DNA (local DNA conformational structure change), especially including those that occur on fast timescales, for example, from femtosecond to nanosecond. The 'snap shot of substrate DNA' revealed can enable us to take a closer look at DNA-protein interactions, facilitating the better understanding of it within high temporal resolution, whose heterogeneity would otherwise be inaccessible by static conventional experimental methods.

In the future, with respect to the time-resolved fluorescence methodology established in this thesis, it would have huge potential for use in relation to the transient unzipping in nucleic acid conformation, especially for those induced by enzyme. This could be useful in understanding how enzymes achieve substrate recognition, binding, catalysis specificity *etc.* Regarding the DNA 3WJ work, 2AP bases can be used to replace adenines in different positions in the helical arms, resulting in different distances between them to the branch point. In this way, time-resolved fluorescence can help to systematically study how the bases at varied distances in 3WJ behave. TseI and EcoP15I could be collectively used to investigate the structure or the numbers of repeats on clinical genome samples containing CAG or CTG trinucleotides. Furthermore, the affinity (dissociation constant, K_d value) of different substrates for EcoP15I, for example EcoP15I with the 2AP-labelled duplexes, fully matched naturally occurring duplexes or mismatched duplexes, could be further explored.

Appendix A---Lectures and Conferences

Postgraduate Lectures and Courses

Inorganic and Biophysical Chemistry Research Areas Seminars
Organic and Bio-Organic Chemistry Research Areas Seminars
School of Chemistry Colloquia
EaStCHEM Academic Paper Writing Workshop
EaStCHEM Thesis Writing Workshop
MTBI (Myers-Briggs Type Indicator)-an Introduction
Postgraduate Course in High Resolution NMR Spectroscopy
Mass Spectrometry: Principles and Practice in the 21st Century
Workshop of the Computational Chemistry Software Package Gaussian09

Conferences and Meetings

The EaStCHEM Organic Chemistry meets Chemical Biology Symposium, University of St Andrews, UK, April 2009

Edinburgh University Chemistry Symposium, Firbush Point, UK, April 2011
Poster: Nanopore-based single-molecule biophysics: probing conformational change of DNA polymerase

Edinburgh University Chemistry Symposium, Firbush Point, UK, May 2011
Presentation: The biology of restriction and anti-restriction

RSC the 7th Nucleic Acid Forum, London UK, July 2011
Poster: Investigating base unzipping by human flap endonuclease using 2-aminopurine as a structural probe

6th Cambridge Symposium on Nucleic Acids Chemistry and Biology, Cambridge UK, September 2011
Poster: TseI restriction enzyme recognises and cuts A:A and T:T mismatches in its target sequence GCWGC

Noreen Murray Symposium: From Genetic Recombination to Recombinant DNA, University of Edinburgh, UK, November 2011

Mass Spectrometry for the Medical & Life Sciences: The Emerging Role of Ion Mobility Mass Spectrometry, The Royal Society of Edinburgh, September 2011

EaStCHEM-GlaxoSmithKline Symposium, University of Edinburgh, March 2012

Appendix B---Reprints of Publications

1. Ma, L.; Cockroft, S. L., Biological Nanopores for Single-Molecule Biophysics. *ChemBioChem* **2010,** *11* (1), 25-34.

2. Wilson, C. P.; Boglio, C.; Ma, L.; Cockroft, S. L.; Webb, S. J., Palladium(II)-Mediated Assembly of Biotinylated Ion Channels. *Chemistry-A European Journal* **2011,** *17* (12), 3465-3473.

3. Sabir, T.; Toulmin, A.; Ma, L.; Jones, A. C.; McGlynn, P.; Schroder, G. F.; Magennis, S. W., Branchpoint Expansion in a Fully Complementary Three-Way DNA Junction. *J. Am. Chem. Soc.* **2012,** 134, 6280-6285.

4. Ma, L.; Chen, K.; Wilson, G. G.; Edwardson, J.; Anne, M.; Jones, A. C.; Dryden, D. T. F., Restriction Endonuclease TseI Cleaves A:A and T:T Mismatches in CAG and CTG Repeats. 2012. (*submitted)*